Glide-related material

1. Richard A. Friesner, Jay L. Banks, Robert B. Murphy, Thomas A. Halgren, Jasna J. Klicic, Daniel T. Mainz, Matthew P. Repasky, Eric H. Knoll, Mee Shelley, Jason K. Perry, David E. Shaw, Perry Francis, and Peter S. Shenkin. Surflex: Fully Automatic Flexible Molecular Docking Using a Molecular Similarity-Based Search Engine. *J. Med. Chem.*, **2004**, *47*, 1739-1749.
2. Thomas A. Halgren, Robert B. Murphy, Richard A. Friesner, Hege S. Beard, Leah L. Frye, W. Thomas Pollard, and Jay L. Banks. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47*, 1750-1759.
3. Richard A. Friesner, Robert B. Murphy, Matthew P. Repasky, Leah L. Frye, Jeremy R. Greenwood, Thomas A. Halgren, Paul C. Sanschagrin, and Daniel T. Mainz. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem.* **2006**, *49*, 6177-6196.
4. Quick start guide.
5. Glide tutorial.

Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy

Richard A. Friesner,*,† Jay L. Banks,‡ Robert B. Murphy,‡ Thomas A. Halgren,‡ Jasna J. Klicic,‡,|| Daniel T. Mainz,‡ Matthew P. Repasky,‡ Eric H. Knoll,† Mee Shelley,§ Jason K. Perry,§ David E. Shaw,# Perry Francis,§ and Peter S. Shenkin‡

Department of Chemistry, Columbia University, New York, New York 10036, Schrödinger, L.L.C., 120 W. 45th Street, New York, New York 10036, Schrödinger, L.L.C., 1500 SW First Avenue, Portland, Oregon 97201, and D. E. Shaw Research and Development, 120 W. 45th Street, New York, New York 10036

Received December 24, 2003

Unlike other methods for docking ligands to the rigid 3D structure of a known protein receptor, Glide approximates a complete systematic search of the conformational, orientational, and positional space of the docked ligand. In this search, an initial rough positioning and scoring phase that dramatically narrows the search space is followed by torsionally flexible energy optimization on an OPLS-AA nonbonded potential grid for a few hundred surviving candidate poses. The very best candidates are further refined via a Monte Carlo sampling of pose conformation; in some cases, this is crucial to obtaining an accurate docked pose. Selection of the best docked pose uses a model energy function that combines empirical and force-field-based terms. Docking accuracy is assessed by redocking ligands from 282 cocrystallized PDB complexes starting from conformationally optimized ligand geometries that bear no memory of the correctly docked pose. Errors in geometry for the top-ranked pose are less than 1 Å in nearly half of the cases and are greater than 2 Å in only about one-third of them. Comparisons to published data on rms deviations show that Glide is nearly twice as accurate as GOLD and more than twice as accurate as FlexX for ligands having up to 20 rotatable bonds. Glide is also found to be more accurate than the recently described Surflex method.

1. Introduction

The number of drug-discovery projects that have a high-resolution crystal structure of the receptor available has increased in recent years and is expected to continue to rise because of the human genome project and high-throughput crystallography efforts. A common computational strategy in such a case is to dock molecules from a physical or virtual database into the receptor and to use a suitable scoring function to evaluate the binding affinity. A number of docking programs are employed extensively in the pharmaceutical and biotechnology industries,[1−10] of which the most widely used appear to be GOLD,[1] FlexX,[2] and DOCK.[3] Over the past several years, considerable success has been reported for these programs in virtual screening applications.[11−13] However, none as of now can be viewed as offering a robust and accurate solution to the docking problem, even in the context of a rigid protein receptor.

In this paper, we describe a new docking methodology that has been implemented in the FirstDiscovery software package Glide[14] (grid-based ligand docking with energetics). Glide has been designed to perform as close to an exhaustive search of the positional, orientational, and conformational space available to the ligand as is feasible while retaining sufficient computational speed to screen large libraries. This has been accomplished via the use of a series of hierarchical filters, as described below. The current performance characteristics of Glide are as follows:

(i) Docking times average less than 1 min for data sets having 0−10 rotatable bonds on an AMD Athelon MP 1800+ processor running Linux.

(ii) Robustness in binding mode prediction is qualitatively superior to what is reported in the current literature for docking methods in widespread use. For example, a comparison with results obtained by the developers of GOLD yields an average rmsd of 1.46 Å for Glide compared with 2.56 Å for GOLD for the 72 noncovalently bound cocrystallized ligands of the GOLD test set[15] that have 10 or fewer rotatable bonds. The comparison to FlexX is even more favorable. Comparisons for ligands having up to 20 rotatable bonds yield similar results.

(iii) Binding affinity predictions, compared with experimental data for cocrystallized complexes, are reasonable (2.3 kcal/mol rmsd), though clearly subject to improvement.

(iv) Results for library screening, reported in the following paper,[16] are very encouraging. Furthermore, database enrichment factors obtained using Glide 2.5 are significantly higher than those obtained using previous versions of Glide.

The paper is organized as follows. Section 2 summarizes the computational methodology used by Glide, while section 3 describes Glide's approach to scoring relative ligand binding affinities. The fourth section

* To whom correspondence should be addressed. Phone: 212-854-7606. Fax: 212-854-7454. E-mail: rich@chem.columbia.edu.
† Columbia University.
‡ Schrödinger, L.L.C., NY.
|| Present address: Serono International, S.A, CH-1211 Geneva 20, Switzerland.
§ Schrödinger, L.L.C., OR.
# Schrödinger, L.L.C., NY and D. E. Shaw Research and Development.

10.1021/jm0306430 CCC: $27.50
Published on Web 02/27/2004

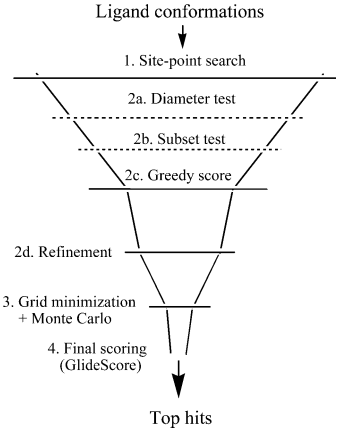

Figure 1. Glide docking "funnel", showing the Glide docking hierarchy.

then presents rmsd values obtained for redocking cocrystallized ligands and compares Glide to GOLD, FlexX, and Surflex, methods we believe to be representative of the current state of the art in high-throughput docking. The fifth section summarizes the results and discusses future directions. Finally, section 6 provides details of the docking methodology and of the optimization of the scoring function; this section also describes the procedure we recommend for protein preparation, which in many cases can substantially affect the quality of the results obtained in docking calculations.

2. Overview of Docking Methodology

Glide uses a series of hierarchical filters to search for possible locations of the ligand in the active-site region of the receptor (Figure 1). The shape and properties of the receptor are represented on a grid by different sets of fields that provide progressively more accurate scoring of the ligand pose. (By "pose" we mean a complete specification of the ligand: position and orientation relative to the receptor, core conformation, and rotamer-group conformations.) These fields are generated as preprocessing steps in the calculation and hence need to be computed only once for each receptor.

The next step produces a set of initial ligand conformations. These conformations are selected from an exhaustive enumeration of the minima in the ligand torsion-angle space and are represented in a compact combinatorial form. Given these ligand conformations, initial screens are performed over the entire phase space available to the ligand to locate promising ligand poses. This prescreening drastically reduces the region of phase space over which computationally expensive energy and gradient evaluations will later be performed while at the same time avoiding the use of stochastic methods; such methods can miss key phase-space regions a certain fraction of the time, thus precluding development of a truly robust algorithm. To our knowledge, Glide is unique in its reliance on the techniques of exhaustive systematic search, though approximations and truncations are required to achieve acceptable computational speed.

Starting from the poses selected by the initial screening, the ligand is minimized in the field of the receptor using a standard molecular mechanics energy function (in this case, that of the OPLS-AA force field[17]) in conjunction with a distance-dependent dielectric model. Finally, the three to six lowest-energy poses obtained in this fashion are subjected to a Monte Carlo procedure that examines nearby torsional minima. This procedure is needed in some cases to properly orient peripheral groups and occasionally alters internal torsion angles.

We and others have found that a conventional molecular mechanics energy function is a reasonable model for predicting binding modes, even in the absence of solvent. However, it is not inadequate for ranking disparate ligands, for example, ligands with different net charge. Therefore, we have implemented a modified and expanded version of the ChemScore[18] scoring function, GlideScore, for use in predicting binding affinity and rank-ordering ligands in database screens. However, we use a combination of GlideScore, the ligand−receptor molecular mechanics interaction energy, and the ligand strain energy to select the correctly docked pose. We find that this composite scoring function, which we call Emodel, is much better at selecting the correct pose than is either the molecular mechanics energy or GlideScore alone.

A final and very important issue is that the scoring function, particularly the molecular mechanics component, needs to be modified in order to accommodate the fact that the protein structure used for docking will not in general be optimized to fit a particular ligand. We have found that the most severe problem when docking a library of ligands into a single rigid receptor structure is the inability of some actives to fit into the protein cavity because the cavity is too small. Therefore, we typically scale down the van der Waals radii of selected (e.g., nonpolar) protein and/or ligand atoms to create additional space in the binding pocket. Our database enrichment studies show that this approach is effective in that context.[16] Better enrichment can often be obtained by tuning the scaling parameters for a given receptor,[16] but the default values are suitable for routine use.

3. Scoring Function

The starting point for Glide scoring is the empirically based ChemScore function of Eldridge et al.,[18] which can be written as

$$\Delta G_{\text{bind}} = C_0 + C_{\text{lipo}}\sum f(r_{\text{lr}}) + C_{\text{hbond}}\sum g(\Delta r)\, h(\Delta\alpha) + C_{\text{metal}}\sum f(r_{\text{lm}}) + C_{\text{rotb}}H_{\text{rotb}} \quad (1)$$

The summation in the second term extends over all ligand-atom/receptor-atom pairs defined by ChemScore as lipophilic, while that in the third term extends over all ligand−receptor hydrogen-bonding interactions. In eq 1, f, g, and h are functions that give a full score (1.00) for distances or angles that lie within nominal limits and a partial score (1.00−0.00) for distances or angles that lie outside those limits but inside larger threshold values. For example, $g(\Delta r)$ is 1.00 if the H···X hydrogen-bond distance is within 0.25 Å of a nominal value of 1.85 Å but tails off to zero in a linear fashion if the distance lies between 2.10 and 2.50 Å. Similarly, $h(\Delta\alpha)$ is 1.00 if the Z−H···X angle is within 30° of 180° and decreases to zero between 150° and 120°.

Glide 2.5 employs two forms of GlideScore: (i) GlideScore 2.5 SP, used by Standard-Precision Glide; (ii) GlideScore 2.5 XP, used by Extra-Precision Glide. These functions use similar terms but are formulated with different objectives in mind. Specifically, GlideScore 2.5 SP is a "softer", more forgiving function that is adept at identifying ligands that have a reasonable propensity to bind, even in cases in which the Glide pose has significant imperfections. This version seeks to minimize false negatives and is appropriate for many database screening applications. In contrast, GlideScore 2.5 XP is a harder function that exacts severe penalties for poses that violate established physical chemistry principles such as that charged and strongly polar groups be adequately exposed to solvent. This version of GlideScore is more adept at minimizing false positives and can be especially useful in lead optimization or other studies in which only a limited number of compounds will be considered experimentally and each computationally identified compound needs to be as high in quality as possible. In what follows, we discuss the development and parametrization of Glide 2.5 SP; XP docking and scoring[19] will be described in a subsequent paper.

GlideScore 2.5 modifies and extends the ChemScore function as follows:

$$\Delta G_{\text{bind}} = C_{\text{lipo-lipo}}\sum f(r_{\text{lr}}) + C_{\text{hbond-neut-neut}}\sum g(\Delta r)\, h(\Delta\alpha) + C_{\text{hbond-neut-charged}}\sum g(\Delta r)\, h(\Delta\alpha) + C_{\text{hbond-charged-charged}}\sum g(\Delta r)\, h(\Delta\alpha) + C_{\text{max-metal-ion}}\sum f(r_{\text{lm}}) + C_{\text{rotb}}H_{\text{rotb}} + C_{\text{polar-phob}}V_{\text{polar-phob}} + C_{\text{coul}}E_{\text{coul}} + C_{\text{vdW}}E_{\text{vdW}} + \text{solvation terms} \quad (2)$$

The lipophilic−lipophilic term is defined as in ChemScore. The hydrogen-bonding term also uses the ChemScore form but is separated into differently weighted components that depend on whether the donor and acceptor are both neutral, one is neutral and the other is charged, or both are charged. In the optimized scoring function, the first of these contributions is found to be the most stabilizing and the last, the charged-charged term, is the least important. The metal−ligand interaction term (the fifth term in eq 2) uses the same functional form as is employed in ChemScore but varies in three principal ways. First, this term considers only interactions with anionic acceptor atoms (such as either of the two oxygens of a carboxylate group). This modification allows Glide to recognize the evident strong preference for coordination of anionic ligand functionality to metal centers in metalloproteases.[20,21] In addition, Glide 2.5 counts just the single best interaction when two or more metal ligations are found. We set the coefficient to −2.0 kcal/mol, a value we believe to be reasonable, though the parameter refinement would have preferred an even more strongly negative value. Third, we assess the net charge on the metal ion in the unligated apo protein (generally straightforward via examination of the directly coordinated protein side chains). If the net charge is positive, the preference for an anionic ligand is incorporated into the scoring function. On the other hand, if the ion is net neutral (as it is, for example, in the case of the zinc metalloprotein farnesyl protein transferase, which accepts neutral ligands such as substituted imidazoles[22]), the preference is suppressed. The seventh term, from Schrödinger's active site mapping facility, rewards instances in which a polar but non-hydrogen-bonding atom (as classified by ChemScore) is found in a hydrophobic region.

Table 1. OPLS-AA Interaction Energies with Full and Modified Charge Distributions for Ionic Centers and Groups ($\epsilon = 2r$)

system	full charges	reduced charges
charged−charged		
$Zn^{2+}\cdots MeOPO_3^{2-}$	−111.1	−16.8
$Zn^{2+}\cdots (MeO)_2PO_2^-$	−70.8	−14.5
$Zn^{2+}\cdots Me_2PO_2^-$	−68.0	−13.4
$Zn^{2+}\cdots$acetate	−80.8	−14.4
$Mg^{2+}\cdots$acetate	−90.0	−16.2
$Mn^{2+}\cdots$acetate	−71.7	−12.6
$Zn^{2+}\cdots CH_3S^-$	−48.7	−11.5
$NH_4^+\cdots$acetate	−24.8	−7.1
Me-guanidinium$\cdots$acetate	−31.6	−11.8
benzamidinium$\cdots$acetate	−27.3	−9.2
$His^+\cdots$acetate	−22.3	−7.4
Me-guanidinium$\cdots (MeO)_2PO_2^-$	−27.7	−12.0
charged−polar		
$Zn^{2+}\cdots H_2O$	−23.8	−10.3
$NH_4^+\cdots H_2O$	−7.2	−5.6
$H_2O\cdots$acetate	−8.6	−4.9
Me-guanidinium$\cdots H_2O$	−8.1	−6.4
benzamidinium$\cdots H_2O$	−6.9	−4.8
$His^+\cdots H_2O$	−7.2	−6.5
$H_2O\cdots Me_2PO_2^-$	−7.2	−4.4
$H_2O\cdots CH_3S^-$	−4.0	−2.4
polar−polar		
$H_2O\cdots H_2O$	−3.7	−3.7

The second major component is the incorporation of contributions from the Coulomb and vdW interaction energies between the ligand and the receptor. To make the gas-phase Coulomb interaction energy a better predictor of binding (and a better contributor to a composite scoring function), we reduce, by ∼50%, the net ionic charge on formally charged groups such as carboxylates and guanidiniums; we also reduce the vdW interaction energies for the atoms directly involved.[23] Table 1 illustrates the effect of these changes for some prototype systems. The wide disparities in the original interaction energies are greatly reduced, though charge−charge interactions are still favored to some extent. The Coulomb−vdW energies used in GlideScore 2.5 (but not those used in Emodel) employ these reductions in net ionic charge except in the case of anionic ligand−metal interactions, for which Glide uses the full interaction energy.

The third major component is the introduction of a solvation model. Like other scoring functions of this type, previous versions of GlideScore did not properly take into account the severe restrictions on possible ligand poses that arise from the requirement that charged and polar groups of both the ligand and protein be adequately solvated. Charged groups, in particular, require very careful assessment of their access to solvent. In addition, water molecules may be trapped in hydrophobic pockets by the ligand, also an unfavorable situation.

To include solvation effects, Glide 2.5 docks explicit waters into the binding site for each energetically

competitive ligand pose and employs empirical scoring terms that measure the exposure of various groups to the explicit waters. This "water-scoring" technology has been made efficient by the use of grid-based algorithms. Using explicit waters, as opposed to a continuum solvation model, has significant advantages. In the highly constrained environment of a protein active site containing a bound ligand, the location and environment of individual water molecules become important. Current continuum solvation models have difficulty capturing these details, but our explicit-water approach has allowed us to develop consistently reliable descriptors for rejecting a high fraction of the false positives that appear in any empirical docking calculation. Our analysis also produced trial values for the coefficients of the various penalty terms. For the most part, these coefficients were used without modification in GlideScore 2.5 XP.[19] For Glide 2.5 SP, however, the need to make the program relatively fast limits the amount of sampling that can be done during docking and hence limits the accuracy of the docked poses. As a result, optimization of the solvation penalties led to smaller coefficients that do not too heavily penalize misdocked actives. Even these smaller values, however, provide substantial enhancement in enrichment factors for many database screens, thrombin being the most prominent example (in this case primarily by rejecting false positives that bury charged groups in a hydrophobic region of the thrombin active site).

4. Docking Accuracy

This section characterizes Glide's performance in reproducing the geometries of cocrystallized ligands taken from an extensive set of 282 publically available PDB[24] complexes. This set includes most of the members of the well-known GOLD and FlexX test sets, approximately 50 PDB complexes used in evaluations of Glide by prospective customers and approximately 50 more complexes whose experimental binding affinities have been used to develop one or more of the empirical scoring functions described in the literature (e.g., ChemScore[18]). We used the latter complexes, and others included in the GOLD and FlexX test sets, to calibrate the GlideScore function. Our coverage of the GOLD and FlexX sets is not quite complete because Glide does not deal with covalently attached ligands (seven cases: 1aec, 1ase, 1blh, 1tpp, 1lmp, 3gch, and 4est) and cannot handle ligands having more than 35 rotatable bonds (one case: 2er6). In addition, we excluded one complex (6rsa) because it contains an atomic species (vanadium) for which the OPLS-AA force field used by Glide has no parameters.

All results were obtained with release 2.5021 of the FirstDiscovery suite[14] on an AMD Athelon MP 1800+ processor running Linux. All structures were prepared using the protein-preparation procedure described in section 6 or an earlier version of that procedure. For these calculations, the vdW radii of nonpolar protein atoms were not scaled, but the radii of nonpolar ligand atoms (taken to be atoms having a partial charge of less than 0.15 e^- in magnitude) were scaled down by a factor of 0.8; the same default scaling is also employed in the database-enrichment studies presented in the accompanying paper.[16]

Table 2. Average rms Deviations for Flexible Docking on 282 PDB Complexes[a]

no. of rotatable bonds	no. of cases	av rms top-ranked pose (Å)	av CPU time (min)
0−3	51	1.01	0.2
4−6	92	1.64	0.6
7−10	48	1.79	1.7
0−8	164	1.48	0.5
0−10	191	1.51	0.8
0−20	263	1.89	2.4

[a] Times are AMD Athelon MP 1800+ CPU minutes.

To examine the dependence of the results on the starting geometries, we docked five sets of ligand geometries. One set consists of the (restraint-optimized) native ligands obtained from the protein-preparation procedure. These ligands typically have an rmsd for nonhydrogen atoms of 0.3 Å or less from the original PDB coordinates. The second set uses MMFF94s-optimized[25] versions of the native ligands. Except for 1cps, 1d8f, 1hbv, 1pro, and 2cht,[26] however, the primary set used for assessing docking accuracy and for comparison to GOLD and FlexX consists of MMFF94s-optimized geometries obtained via a short MacroModel conformational search, starting from MMFF94s-optimized versions of the (restraint-optimized) native-ligand geometries. In each case, the optimizations used a 4*r* distance-dependent dielectric model. The fourth set consists of ligand geometries obtained using Corina,[27] while the fifth set contains geometries obtained by optimizing the Corina structures with MMFF94s.

We emphasize that the initial geometry of the ligand never explicitly enters as a docked conformation. However, the conformations sampled by Glide's conformation generator depend on the input bond lengths and bond angles because these variables are not optimized. Furthermore, the fact that the potential-energy landscape has multiple minima and that finite sampling is done by Glide means that different solutions can be obtained from different starting points, even when they are very close to one another. There is also a dependence on input ring conformation, though Glide by default generates and docks alternative conformations of saturated or partly saturated five- and six-membered rings when it deems them to be energetically accessible. We generated the conformationally optimized ligand geometries cited above to make sure that no "memory" of the cocrystallized pose influenced the docking results.

All results are for flexible dockings carried out using Glide's internal conformation generator. With the exception of terminal CH_3, NH_3^+, and NH_2 groups, all rotatable bonds are treated as optimization variables. In addition, as noted above, alternative ring conformations are considered for saturated or partly saturated five- and six-membered rings, and inversion at nitrogen is performed for asymmetric trigonal nitrogen centers in compounds such as sulfonamides. The reported rms deviations in coordinate positions are based on heavy (non-hydrogen) atoms, as is also done by GOLD[1] and FlexX,[28] and are computed relative to the coordinates of the restraint-optimized native-ligand structures obtained from the protein-preparation procedure.

Table 2 summarizes the rms errors obtained as a function of the flexibility of the ligand. As expected, rms errors and CPU times increase with ligand flexibility.

Table 3. Comparison of rms Deviations (Å) for Flexible Docking by Glide and GOLD

	≤10 rotatable bonds (72 cases)		≤20 rotatable bonds (86 cases)		all ligands (93 cases)	
method	av rmsd	max rmsd	av rmsd	max rmsd	av rmsd	max rms
Glide	1.46	8.5	1.65	8.5	1.85	13.7
GOLD	2.56	14.0	2.92	14.0	3.06	14.0

Table 4. Comparison of rms Deviations (Å) for Flexible Docking by Glide and FlexX

	≤10 rotable bonds (133 cases)		≤20 rotatable bonds (175 cases)		all ligands (189 cases)	
method	av rmsd	max rmsd	av rmsd	max rmsd	av rmsd	max rms
Glide	1.38	8.5	1.70	9.2	1.95	13.7
FlexX	2.99	12.6	3.48	13.4	3.72	15.5

Both, however, are quite modest for sets of ligands having 0−8 or 0−10 rotatable bonds such as are often employed in database screens carried out to find new leads. In general, the docking performance of Glide is very reasonable over a wide range of rotatable bonds and chemical functionality. Detailed results are given in Table S1 (Supporting Information); these results show that Glide reproduces the experimentally measured binding affinities for 128 cocrystallized ligands with an rms deviation of 2.3 kcal/mol and produces rms deviations from the cocrystallized ligand position that are less than 1 Å for nearly half of the 282 cases and are greater than 2 Å in only about one-third of the cases.

Comparison to GOLD and FlexX. Detailed docking accuracy results for GOLD and FlexX are posted on the GOLD[15] and FlexX[29] web sites. These data have enabled us to make the head-to-head comparisons shown in the tables below.

Tables 3 and 4 compare rms deviations (Å) given by Glide and GOLD and by Glide and FlexX for common sets of noncovalently bound ligands having up to 10 and up to 20 rotatable bonds as well as for all ligands Glide can handle (i.e., up to 35 rotatable bonds). In some cases, the PDB structure available when the GOLD or FlexX work was done is no longer accessible but a later structure is available. In such cases, we have used the later submission. For example, Glide uses 4aah whereas GOLD and FlexX use 3aah. The Glide calculations use the conformationally optimized versions of the native ligands. The comparison for ligands having 10 or fewer rotatable bonds seems to us the most relevant to database screening applications, which usually seek to find leads that are relatively inflexible. On average, Glide gives rms deviations that are less than 60% of those given by GOLD and less than half those given by FlexX. The comparison for ligands having up to 20 rotatable bonds is also favorable.

Table 5 presents the detailed results on which the summaries in Tables 3 and 4 are based. This listing shows that Glide gives a better result nearly twice as often as GOLD and more than 4 times as often as FlexX. While far from perfect, we believe that this performance represents a qualitative improvement in docking accuracy. We should note, however, that these comparisons may not be completely fair to GOLD and FlexX. The ligands were prepared in a comparable manner for all three methods, i.e., by subjecting the native ligand structures to a force-field-based optimization procedure before docking,[1,28] though Glide also used conformational search to ensure that the starting ligand structures had no memory of the cocrystallized pose. A difference arises in the preparation of the protein sites, however, in that both the GOLD and FlexX calculations retained the original PDB coordinates for non-hydrogen atoms. It is possible that these methods might give more accurate dockings if they used our protein and ligand preparations, in which steric clashes have been annealed away. However, while the GOLD study cites four cases of noncovalently bound ligands in which the crystallographic ligand geometry appears to be incorrect (1apt, 1tdb, 1hef, and 1ive),[1] it does not ascribe any of the problematic dockings to steric clashes between the ligand and the protein; for FlexX, only one such instance (1srj) is cited.[28]

One reason for suspecting that GOLD and FlexX may be less sensitive to the details of the protein preparation is that neither uses the hard 12-6 Lennard-Jones vdW potential employed by Glide. GOLD does employ an 8-4 potential,[1] but this potential is much more forgiving of nonbonded incursions. For example, it penalizes a contact at 70% of the sum of the vdW radii by only 0.9−1.8 kcal/mol, whereas Glide penalizes such a contact by 5.5−11 kcal/mol (assuming a well depth of −0.1 to −0.2 kcal/mol). FlexX does not use a molecular mechanics expression for nonbonded repulsion. Indeed, it may use no repulsive function at all, though it does reject dockings for which the "overlap volume" exceeds 2.5 Å^3 for any particular pair of ligand and protein atoms or 1.0 Å^3 averaged over all such interactions.[28] In view of these differences, it is not clear that the results for GOLD and FlexX would be materially improved by using our preparations. However, only explicit comparisons that use identical protein preparations can resolve this question.

Comparison to Surflex. Docking accuracy is also better than that recently reported by Jain for Surflex.[30] In particular, for a common set of 78 cocrystallized PDB complexes (of 81 considered by Jain), Glide gives an average rmsd from the cocrystallized ligand pose of 1.35 Å whereas Surflex gives an average rmsd of 1.82 Å. Moreover, Glide places 47 of the 78 ligands wthin 1 Å rmsd as against 38 for Surflex and makes errors of more than 2 Å in 14 cases as against 19 for Surflex.

Cross-Docking for Thymidine Kinase. Table 6 shows rms deviations to the cocrystallized pose for docking of thymidine kinase actives to the 1kim site by GOLD, FlexX, and DOCK,[12] by Surflex,[30] and by Glide. All methods have trouble docking the acv, gcv, and pcv ligands. This is expected because acv, gcv, and pcv are purine-based ligands that do not fit properly into the pyrimidine-based 1kim site. The reason is that the Gln 125 side chain undergoes a 180° rotation on going from a pyrimidine site to a purine site and the geometry that is correct for the parent site has an acceptor−acceptor and/or a donor−donor clash in the alternative site. For the seven pyrimidine-based ligands, Glide does very well except for hmtt, which does not fit quite properly into the 1kim site when the nonpolar ligand vdW radii are scaled by 0.8 (the default). Surflex and GOLD also give

Table 5. The rms Deviations (Å) for Glide, GOLD, and FlexX for Members of the GOLD and FlexX Test Sets

complex	Glide	GOLD	FlexX	complex	Glide	GOLD	FlexX	complex	Glide	GOLD	FlexX	complex	Glide	GOLD	FlexX
121p	1.57	n/a	1.29	1aaq	1.30	12.85	1.75	1phf	1.14	n/a	4.23	1phg	4.32	1.35	4.74
1abe	0.17	0.86	1.16	1abf	0.20	n/a	1.27	1poc	5.09	1.27	9.25	1ppc	7.92	n/a	3.05
1acj	0.28	4.00	0.49	1acm	0.29	0.81	1.39	1pph	4.31	n/a	4.91	1ppi	6.24	n/a	6.91
1aco	1.02	0.86	0.96	1aha	0.11	0.51	0.56	1ppk	0.45	n/a	1.54	1ppl	2.82	n/a	5.62
1ake	3.35	n/a	1.18	1apt	0.58	1.62	1.89	1ppm	0.62	n/a	8.27	1pso	13.10	n/a	1.61
1atl	0.94	n/a	2.06	1avd	0.52	n/a	1.22	1rbp	0.96	n/a	1.13	1rds	3.75	4.78	4.89
1azm	1.87	2.52	2.37	1baf	0.76	6.12	8.27	1rne	10.08	2.00	12.24	1rnt	0.72	n/a	1.90
1bbp	4.96	n/a	3.75	1bma	9.31	n/a	13.41	1rob	1.85	3.75	7.70	1slt	0.51	0.78	1.63
1byb	10.49	n/a	1.62	1cbs	1.96	n/a	1.68	1snc	1.91	n/a	7.48	1srj	0.58	0.42	2.36
1cbx	0.36	0.54	1.35	1cde	1.29	n/a	7.45	1stp	0.59	0.69	0.65	1tdb	1.46	10.48	10.10
1cdg	3.98	n/a	4.87	1cil	3.82	n/a	3.85	1thy	2.31	n/a	2.67	1tka	2.28	1.88	1.17
1com	3.64	n/a	1.62	1coy	0.28	0.86	1.06	1tlp	1.86	n/a	2.85	1tmn	2.80	1.68	0.86
1cps	3.00	0.84	0.99	1ctr	3.56	n/a	2.82	1tng	0.19	n/a	1.93	1tnh	0.33	n/a	0.56
1dbb	0.41	1.17	0.81	1dbj	0.20	0.72	1.22	1tni	2.18	n/a	2.71	1tnj	0.35	n/a	0.89
1dbk	0.47	n/a	0.76	1dbm	1.97	n/a	2.08	1tnk	0.87	n/a	1.41	1tnl	0.23	n/a	0.71
1did	3.82	3.72	4.22	1die	0.79	1.03	4.71	1tph	0.20	n/a	1.50	1tpp	1.12	0.43	1.11
1dr1	1.47	1.41	5.64	1dwb	0.25	n/a	0.54	1trk	1.64	n/a	1.57	1tyl	1.06	n/a	2.34
1dwc	0.87	n/a	1.19	1dwd	1.32	1.71	1.66	1ukz	0.37	n/a	0.94	1ulb	0.28	0.32	3.37
1eap	2.32	3.00	3.72	1eed	5.90	12.43	9.78	1wap	0.12	n/a	0.57	1xid	4.30	0.92	2.01
1ela	1.60	n/a	9.71	1elb	4.40	n/a	7.17	1xie	3.86	0.69	1.94	2ack	0.97	4.99	2.21
1elc	8.22	n/a	4.74	1eld	0.67	n/a	6.98	2ada	0.53	0.40	0.67	2ak3	0.71	5.08	0.91
1ele	2.52	n/a	10.73	1epb	1.78	2.08	2.77	2cgr	0.38	0.99	3.53	2cht	0.42	0.59	4.58
1eta	2.92	11.21	8.46	1etr	1.48	4.23	7.24	2cmd	0.65	n/a	3.75	2cpp	0.17	n/a	2.94
1fen	0.66	n/a	1.39	1fkg	1.25	1.81	7.59	2ctc	1.61	0.32	1.97	2dbl	0.69	1.31	1.49
1fki	1.92	0.71	0.59	1frp	0.27	n/a	1.89	2gbp	0.15	n/a	0.92	2lgs	7.55	n/a	4.63
1ghb	1.89	1.45	1.33	1glp	0.34	n/a	0.47	2mcp	1.30	4.37	2.07	2phh	0.38	0.72	0.43
1glq	0.29	1.35	6.43	1hdc	0.58	10.49	11.74	2pk4	0.86	1.34	1.66	2plv	1.88	13.92	7.85
1hdy	1.74	0.94	n/a	1hef	5.30	1.87	15.32	2r04	0.80	n/a	12.55	2r07	0.48	8.23	11.63
1hfc	2.24	n/a	2.51	1hgg	2.10	n/a	10.05	2sim	0.92	0.92	1.99	2tmn	0.58	n/a	5.16
1hgh	0.28	n/a	4.14	1hgi	0.28	n/a	0.97	2xis	0.85	n/a	1.54	2yhx	3.84	1.19	2.25
1hgj	0.18	n/a	3.98	1hri	1.59	14.01	10.23	2ypi	0.31	n/a	1.22	3cla	8.51	5.45	6.42
1hsl	1.31	0.97	0.59	1hti	4.40	n/a	1.54	3cpa	2.40	1.58	2.53	3hvt	0.77	1.12	10.26
1hvr	1.50	n/a	3.35	1hyt	0.28	1.10	1.62	3mth	5.48	10.12	1.59	3ptb	0.27	0.96	0.55
1icn	2.34	8.63	10.52	1ida	11.88	12.12	11.95	3tpi	0.49	0.80	1.07	4aah	0.30	0.42	5.93
1igj	1.30	9.42	7.17	1imb	0.89	n/a	4.71	4cts	0.19	1.57	1.53	4dfr	1.12	1.44	1.40
1ivb	4.97	n/a	1.29	1ivc	1.94	n/a	2.21	4fab	4.50	5.69	4.95	4fbp	0.56	n/a	1.78
1ivd	0.72	n/a	5.42	1ive	2.61	2.16	5.34	4fxn	0.44	n/a	1.04	4hmg	0.78	n/a	5.74
1ivf	0.53	n/a	6.97	1lah	0.13	n/a	0.28	4phv	0.38	1.11	1.12	4tim	1.32	n/a	4.09
1lcp	1.98	n/a	1.65	1ldm	0.30	1.00	0.74	4tln	2.24	n/a	3.68	4tmn	1.87	n/a	8.35
1lic	4.87	10.78	5.07	1lmo	0.93	n/a	4.49	4ts1	0.85	n/a	1.41	5abp	0.21	n/a	1.17
1lna	0.95	n/a	5.40	1lst	0.14	0.87	0.71	5cpp	0.59	n/a	1.49	5cts	0.28	n/a	11.61
1mbi	1.68	n/a	0.47	1mcr	4.33	6.23	10.04	5p2p	1.82	1.55	1.00	5tim	0.58	n/a	1.99
1mdr	0.52	0.36	0.88	1mld	0.32	n/a	1.45	5tmn	2.43	n/a	4.38	6abp	0.40	1.08	1.12
1mmq	0.92	n/a	0.52	1mrg	0.30	n/a	0.81	6cpa	4.58	n/a	6.61	6rnt	2.22	1.20	4.79
1mrk	1.20	1.01	3.55	1mup	4.37	3.96	3.82	6tim	1.73	n/a	1.60	6tmn	2.66	n/a	5.10
1nco	6.99	n/a	5.85	1nis	0.97	4.29	1.41	7cpa	4.14	n/a	9.11	7tim	0.14	0.78	1.49
1nsc	1.21	n/a	2.12	1pbd	0.21	0.57	0.33	8atc	0.37	n/a	0.62	8gch	0.30	0.86	8.91
1pha	0.69	1.24	n/a	1phd	1.22	0.85	0.65	9hvp	2.68	n/a	15.54				

Table 6. Accuracy in Cross-Docking of Thymidine-Kinase Inhibitors to the 1kim Site[a]

	rmsd (Å) of best-scoring pose[b]				
ligand	Glide	DOCK	FlexX	GOLD	Surflex
dT	0.45	0.82	0.78	0.72	0.74
ahiu	0.54	1.16	0.88	0.63	0.87
mct	0.79	7.56	1.11	1.19	0.87
dhbt	0.68	2.02	3.65	0.93	0.96
idu	0.35	9.33	1.03	0.77	1.05
hmtt	2.83	9.62	13.30	2.33	1.78
hpt	1.58	1.02	4.18	0.49	1.90
acv	4.22	3.08	2.71	2.74	3.51
gcv	3.19	3.01	6.07	3.11	3.54
pcv	3.53	4.10	5.96	3.01	3.84

[a] The last three ligands are purines that are not expected to fit properly into a pyrimidine-based site such as 1kim (see text). [b] Data for DOCK, FlexX, and GOLD are taken from Rognan and co-workers;[12] data for Surflex are taken from Jain.[30]

quite reasonable results in this case, but FlexX and DOCK fare noticeably more poorly.

Influence of Input Ligand Geometry on Docking Accuracy. Table 7 examines the effect on the final docked structures of using different starting ligand geometries. The native ligands perform best, followed by the conformationally optimized ligands. The MMFF94s-optimized geometries used for the native ligands and the Corina structures do slightly less well but are docked more accurately than are the raw Corina structures. The last comparison shows that it helps to preminimize Corina-derived ligands with a force field such as MMFF94s.

Table 7. Effect of Input Ligand Geometry on Docking Accuracy for Glide Using the GOLD Test Set

ligand set	av rms (Å)
native ligands	1.49
MMFF94s-optimized native ligands	2.00
MMFF94s-conformational search	1.85
Corina	2.35
MMFF94s-optimized Corina	2.02

MMFF94s preminimization can be performed fairly easily on a large database with the premin script provided with Glide. These minimizations employ a 4*r* distance-dependent dielectric model and use MacroMod-

el's efficient truncated Newton minimizer. Processing times are a few tenths of a second per ligand on an AMD Athelon MP 1800+ processor running Linux. The procedure keeps track of instances in which the structure recognition (e.g., atom typing) or minimization fails and automatically submits a series of MacroModel jobs that ultimately collect the good (minimized) and bad (untreatable) structures. The user can examine and fix the bad structures or can discard them.

5. Discussion and Conclusions

This paper has described the new computational algorithms for docking and scoring we have developed for Glide and has evaluated the performance of these algorithms in predicting binding modes over a wide range of cocrystallized structures. While significant errors in binding mode prediction are found in some test cases, robustness has clearly been qualitatively improved compared to widely used alternative packages such as GOLD and FlexX. Docking accuracy is also better than that reported for the recently introduced Surflex method.[30]

Further improvements can be made to Glide with regard to both accuracy and computational efficiency. The algorithms for initial screening and energy minimization are not yet fully optimized, and a 2- to 3-fold reduction in computational effort per ligand may be attainable. The ability to impose constraints on the ligand position (e.g., by requiring that a suitable ligand atom be hydrogen-bonded to a particular protein residue or be coordinated to a metal atom in the protein) is included in Glide 2.5. This approach allows the user to guarantee satisfaction of targeted protein interactions and speeds up the calculations by reducing the size of the phase space that needs to be examined. We expect to continue to improve the scoring functions used in all three phases of the calculation (initial screening, energy minimization, binding affinity prediction). Most importantly, we have developed new sampling and scoring algorithms for Extra-Precision Glide that are proving to be highly efficient at finding correctly docked poses and at rejecting false positives in database screens;[19] these efforts will be described in a subsequent paper.

6. Methods

Conformation Generation. As a first step in its docking protocol, Glide carries out an exhaustive conformational search, augmented by a heuristic screen that rapidly eliminates conformations deemed not to be suitable for binding to a receptor, such as conformations that have long-range internal hydrogen bonds. This procedure eliminates high-energy conformers by evaluating the torsional energy of the various minima using a truncated version of the OPLS-AA molecular mechanics potential function and by imposing a cutoff of the allowed value of the total conformational energy compared to the lowest-energy state. The parameters of the heuristic screening function were optimized via extensive testing on cocrystallized PDB complexes, as described below.

Each ligand is divided into a "core" region and some number of "rotamer groups" (Figure 2), each of which is attached to the core by a rotatable bond but does not itself contain additional rotatable bonds. That is, the core is what remains when each terminus of the ligand is severed at the "last" rotatable bond, as is indicated in the figure (the directly attached atom of each rotomer group is also considered to be part of the core). Carbon and nitrogen end groups terminated with hydrogen ($-CH_3$, $-NH_2$, $-NH_3^+$) are not considered rotatable because their conformational variation is of little interest. Each core region is represented by a set of core conformations, the number of which depends on the numbers of rotatable bonds, conformationally labile five- and six-membered rings, and asymmetric pyramidal trigonal nitrogen centers in the core. As Table 8 shows, this set typically contains fewer than 500 core conformations, even for quite large and flexible ligands, and far fewer for more rigid ligands. Every rotamer state for each rotamer group is enumerated, and the core plus all possible rotamer-group conformations is docked as a single object in Glide. Because each core typically has many rotamer-group combinations, the effective number of conformations being docked can easily number in the thousands or tens of thousands for molecules having several rotatable bonds.

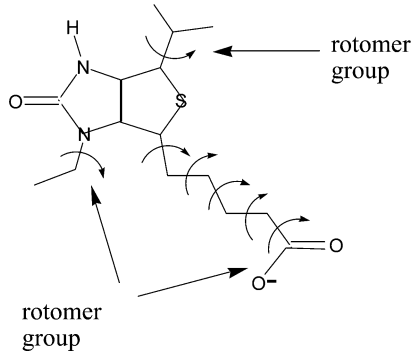

Figure 2. Definition of core and rotamer groups. The four central torsions are part of the core. Note that methyl groups are not considered rotatable.

Table 8. Maximum Number of Conformers Allowed vs Number of Core Degrees of Freedom

no. of core degrees of freedom[a]	max no. of conformers allowed	actual no. of core conformers kept for individual ligands
3	120	4, 13, 17, 18
4	120	4, 6, 18, 19, 24, 34
5	120	9, 106, 106
6	150	24, 28, 150
7	214	214, 214, ...
8	278	278, 278, ...
9	342	342, 342, ...
10	406	406, 406, ...
12	534	534, 534, ...
14	662	662, 662, ...

[a] Number of rotatable bonds in the core plus number of conformationally labile five- and six-membered rings and number of asymmetric trigonal nitrogen centers in compounds such as sulfonamides.

A key issue is how closely one of the conformations matches the correct cocrystallized conformation. An exact match is not needed because the ligand subsequently undergoes flexible torsional optimization. We have found, however, that for best results one of the starting conformations needs to be within about 1.5 Å rmsd of the correct cocrystallized conformation. We have tested our algorithm by applying it to 796 cocrystallized ligands taken from the PDB and by locating the core

Table 9. The rms Deviations between Best-Generated and Cocrystallized Ligands, Expressed as a Percentage of Ligands Falling into Specified rms Distance Ranges

		percent of ligands having an rms deviation in the listed range, Å					
no. of rotatable bonds	no. of ligands	0.00−0.49	0.50−0.99	1.00−1.49	1.50−1.99	2.00−2.99	>3.00
1−10	495	33	37	23	15	2	0
11−15	159	3	29	45	17	6	0
16−20	87	0	10	38	31	20	1
>20	55	0	9	25	29	27	9

plus rotamer-group conformation that best matches the cocrystallized pose. The results, summarized in Table 9, show that only 7% of ligands having 1−10 rotatable bonds have rms deviations between the best core/rotamer-group conformation and the cocrystallized pose of 1.5 Å or greater. For more flexible ligands, the errors are understandably larger. Nevertheless, 77% of ligands having 11−15 rotatable bonds and 48% of ligands having 16−20 rotatable bonds have rms deviations of less than 1.5 Å. Furthermore, 89% of ligands with 11−20 rotatable bonds fall within 2 Å.

Initial Screening of Ligand Poses. For each core conformation (or for rigid docking, each ligand), Glide performs an exhaustive search of possible positions and orientations over the active site of the protein. The search begins with the selection of "site points" on an equally spaced 2 Å grid that permeates the active-site region (step 1 in Figure 1). To make this selection, precomputed distances from the site point to the receptor surface, evaluated at a series of prespecified directions and binned in 1 Å ranges, are compared to binned distances from the ligand center to the ligand surface. For flexible docking, the ligand center is defined as the midpoint of the two most widely separated atoms in the core region (which includes the directly attached atom of each rotamer group); for rigid docking, the two most widely separated atoms in the entire ligand are used. The line through these atoms is called the "ligand diameter." Glide positions the ligand center at the site point if there is a good enough match of the histograms of binned distances but skips over the site point if there is not. The ligand center can be placed at any site point on or within a box (by default, 12 Å on a side) that contains the candidate site points. The generous size of this "ligand center" box ensures that the placement of the docked ligand is not overly constrained.

The second stage examines the placement of atoms that lie within a specified distance of the ligand-diameter axis for a prespecified selection of possible orientations of the ligand diameter (step 2a). If there are too many steric clashes with the receptor, the orientation is skipped. Next (step 2b), the ligand is rotated about the ligand diameter and the subset consisting of the atoms capable of making hydrogen bonds or ligand−metal interactions with the receptor is scored ("subset test"). If this score is good enough, all interactions with the receptor are scored (step 2c).

The scoring in these three tests is carried out using a discretized version of ChemScore[18] in which precomputed scores for the ChemScore atom types are assigned to 1 $Å^3$ boxes. Much as for ChemScore itself, this algorithm recognizes favorable hydrophobic, hydrogen-bonding, and metal-ligation interactions and also penalizes steric clashes. This stage is called "greedy scoring" because the actual score for each atom depends not only on its position relative to the receptor but also on the best possible score it could get by moving ±1 Å in *X*, *Y*, and/or *Z*. This is done to mute the sting of the large 2 Å jumps in the site-point/ligand-center positions. The final step in stage 2 is to rescore the top greedy-scoring poses (typically ∼5000 in number) via a "refinement" procedure (step 2d) in which the ligand as a whole is allowed to move rigidly by ±1 Å in the Cartesian directions.

Energy Minimization Using a Molecular Mechanics Scoring Function. Only a small number of the best refined poses (typically 400) are minimized on precomputed OPLS-AA van der Waals and electrostatic grids for the receptor. The energy and gradient calculations are performed using standard three-dimensional interpolation methods. The Coulomb and van der Waals fields of the protein are stored at the vertexes of a grid, and the interaction of each ligand atom with these fields is evaluated using trilinear interpolation formulas for a cube. Methods of this type are in common use and have been described extensively in the literature, so we do not discuss the mathematical details here.

To ensure sufficient accuracy in regions in which the ligand and protein come into contact, Glide uses a multigrid strategy. The Coulomb/van der Waals grid is initially built using large boxes, typically 3.2 Å on a side, and is then refined hierarchically into boxes of 1.6, 0.8, or 0.4 Å depending on the distance of the box to the van der Waals surface of the protein; the smaller this distance, the higher the resolution of the box used. The resulting mesh tiles the docking volume with cubes of various size. Once the identity of the cube in which a ligand atom is contained is identified, the appropriately scaled interpolation formula can be used. While some extra bookkeeping is required, the additional computational cost as compared to a uniform mesh is negligible, whereas the reduction in memory can be more than 2 orders of magnitude.

The energy minimization typically begins on a set of Coulomb and vdW grids that have been "smoothed" to reduce the large energy and gradient terms that result from too-close interatomic contacts; it finishes on the full-scale OPLS-AA nonbonded energy surface ("annealing"). This minimization consists only of rigid-body translations and rotations when external conformations are docked rigidly. For flexible docking, however, the minimization also includes torsional motion about the core and end-group rotatable bonds. Finally, the three to six lowest-energy poses obtained in this fashion are subjected to the Monte Carlo procedure cited in section 2.

The minimized poses are then rescored using the scoring function described in section 3. As previously noted, the choice of the best-docked structure is made using a model energy score ("Emodel") that combines the energy-grid score, the binding affinity predicted by

GlideScore, and (for flexible docking) the internal strain energy for the model potential used to direct the conformational-search algorithm.

Optimization of the Scoring Function. The parameter optimization for GlideScore 2.5 used a simulated-annealing algorithm that has proven to be very efficient at producing large changes to the input values for the parameters when large changes are warranted. In addition to the terms described in section 3, the fitting process also considered, but ultimately rejected, a number of other prospective terms, including nearly all of those employed in ScreenScore or in its FlexX or PLP predecessors.[13]

The largest component of the objective function in the optimization process was the ranking of active compounds in a large and diverse suite of database screening tests, as measured by enrichment factors computed as described in the following paper.[16] The second component was the fit of the predicted binding affinities for a set of 125 PDB cocrystallized complexes to experimentally measured values; the rmsd achieved in this case is 2.2 kcal/mol, a reasonable result but one that can clearly be improved.

In the parametrization of GlideScore 1.8 and 2.0, we used decoy ligands assembled from cocrystallized PDB complexes and from a small portion of the Comprehensive Medicinal Chemsitry database[31] in the database screens. However, it became apparent through more extensive testing that these ligands, which average only 290 in molecular weight, are too small to make enough favorable interactions with the protein site to compete fairly with the known actives for our screens, which average about 410 in molecular weight when the very large HIV protease ligands are excluded. This disparity in molecular weight distorted both the apparent enrichment factors in the database screens and the parameters we obtained from the optimization process.

A second set of database ligands used in the present work was provided recently by a pharmaceutical colleague. This set consists of known drugs and other compounds identified in drug-discovery projects. Because our colleage intentionally chose the compounds to be "relatively small" (their average molecular weight is 337), they, too, are smaller than typical drugs and investigational compounds, which judging from Oprea's survey of the MDDR database average about 410 or 420 in molecular weight.[32] A third set of average molecular weight 350 (the "pc-350" set) consists of 1000 representative compounds drawn from a ~1 million compound database of purchasable compounds recently assembled by Schrödinger. Finally, two sets of 1000 "druglike" compounds of average molecular weight 360 and 400, the "dl-360" and "dl-400" sets, were also drawn from the million-compound database. For the pc and dl datasets, neutral database compounds were first modified by FirstDiscovery's ionizer utility to protonate or deprotonate ionizable functional groups (subject to limits of ±2 on the net charge and to a total of no more than four charged groups) to yield ionic states likely to be present in measurable concentration between pH 5 and 9. (This is to allow for shifts in pK_a induced by the protein site.) The dl sets were selected to mimic the property distribution of the drug/lead set by using a precursor to the FirstDiscovery ligparse facility. The FirstDiscovery premin utility was then employed to minimize the pc and dl ligands with MMFF94s,[25] using a "4*r*" distance-dependent dielectric. All compounds considered had 100 or fewer atoms and 20 or fewer rotatable bonds.

Table 10. Properties of 1000-Compound Ligand Databases Used in Database Screens[a]

property (av)	cmc/pdb[b]	drug/lead[c]	pc-350[d]	dl-360[e]	dl-400[e]
molecular weight	290	337	350	360	400
atoms	38.30	42.83	42.79	45.74	50.75
non-hydrogen atoms	22.44	26.37	25.63	28.13	31.28
higher-row atoms	0.52	0.47	0.82	0.50	0.56
% hydrophobic carbons	58.52	59.12	63.67	58.94	59.12
rings	1.53	2.09	2.33	2.23	2.51
heteroaramatic rings	0.24	0.84	0.40	0.90	0.99
rotatable bonds	5.51	5.80	5.95	6.19	6.92
amide hydrogens	0.39	0.27	0.66	0.29	0.33
neutral donors	1.61	1.46	1.02	1.56	1.70
charged donors	0.88	0.96	0.30	1.02	1.14
neutral acceptors	1.87	1.49	2.09	1.59	1.78
charged acceptors	0.63	0.42	0.11	0.45	0.49
divalent oxygens	0.55	0.68	0.78	0.73	0.82
neutral amines	0.05	0.01	0.02	0.11	0.13
acidic hydrogens	0.02	0.01	0.00	0.04	0.04

[a] Each ligand database contains 1000 compounds. [b] Taken from a subset of the Comprehensive Medicinal Chemistry database or from cocrystallized PDB complexes. [c] Compounds of relatively low molecular weight from the Derwent World Drug Index provided by a pharmaceutical collaborator. [d] Extracted from a 1-million-compound database of purchasable compounds assembled by Schrödinger in such a way as to preserve the distribution of listed properties. [e] Extracted from the million-compound database in such a way as to preserve proportionately scaled values of properties of the "drug/lead" set; however, the distribution of the last two listed properties was not controlled for.

The properties of these ligand databases are shown in Table 10. We believe that the dl-400 set is representative of ligands one would expect to find in the compound collection of a pharmaceutical or biotechnology company.

We weighted the dl-400 set the most heavily in the parametrization of GlideScore 2.5, but to broaden the parametrization, we used the others as well and also included screens run with nonstandard values for the protein and ligand scaling factors. The large number of screens employed (94 in all, embracing 16 receptor sites) should guard against overfitting and help to make database screening with Glide as tolerant as possible to variations in the ligand sets and the vdW scale factors.

The lipophilic-contact term in Glide 2.5 SP scoring (eq 2) contributes −4.85 kcal/mol to an average score of −9.33 kcal/mol for active compounds included in the database screens, and a second contact term, the vdW interaction energy, contributes −2.39 kcal/mol on average. Thus, these two terms account for nearly 80% of the total score. The hydrogen-bonding terms are next largest in importance, their average contributions for the actives being −1.12 kcal/mol, representing roughly two hydrogen bonds. One major change in the scoring concerns the way in which hydrogen bonding is evaluated. While GlideScore 2.0 differentiated hydrogen bonds on the basis of charge, our investigation of database screening results and PDB cocrystallized structures led us to a new understanding of how hydrogen bonds can be further differentiated. We believe the contribution of hydrogen bonds to binding

affinity depends substantially on the details of the hydrogen bonding in certain specific ways. On the basis of these insights, we programmed terms beyond those shown in eq 2 into Glide and optimized them against our entire range of database screens. While this is still an ongoing research area, the current parametrization leads to a substantial improvement compared to the treatment in Glide 2.0. One similarity, however, is that neutral−neutral hydrogen bonds make the largest contribution, followed by charged−neutral and then charged−charged hydrogen bonds.

Protein Preparation. Our philosophy in the application of rigid docking methods to virtual screening is that, if possible, information based on existing cocrystallized structures for the receptor of interest should be exploited. Obviously, there will be situations in which no experimentally determined structure exists, e.g., when dealing with genomic targets. In this case, a variety of strategies are possible, for example, the use of homology modeling based on cocrystallized structures for related proteins. We focus here, however, on the case in which an experimental structure is available.

Our procedure normally starts with a protein and a cocrystallized ligand. It finishes with a partially optimized protein−ligand complex to which hydrogens have been added, subject to adjustment of protonation states for ionizable residues, modification of tautomeric forms for histidine residues, and repositioning of reorientable hydrogens (e.g., side chain hydroxyl hydrogens).

The first step is to prepare the cocrystallized ligand by making sure that multiple bonds are defined correctly and that hydrogens are properly added. Normally, proteins are provided without attached hydrogens. When hydrogens are present, we usually delete all except those in peptide bonds. In such a case, it is important to note where the authors of the structure have assigned nonstandard protonation states so that these can be reimposed later if this appears warranted. Cofactors, which are included as part of the protein, need to have multiple bonds and formal charges assigned properly so that hydrogens will be added correctly in a later step.

The second step uses the pprep script that is provided with Glide. This procedure neutralizes residues that do not participate in salt bridges and that are more than a specified distance from the nearest ligand atom.[33] By default, it choses the value between 10 and 20 Å for this distance that minimizes the total charge of the receptor plus ligand complex. The script also sets the tautomeric state for His residues, which are assumed to be neutral, by considering potential metal-ligation and hydrogen-bonding interactions.

The third step is to postprocess the pprep'd receptor. This is necessary because the judgments made by the preparation procedure will not always be correct. For example, for an aspartyl protease such as HIV, both active-site aspartates will be close to one or more atoms of a properly docked ligand and hence will be represented as negatively charged. One of these aspartates, however, is typically taken to be neutral in modeling studies. Similarly, Glu 143 in thermolysin typically interacts with an acceptor oxygen of the bound ligand. This residue may need to be protonated, as may His 231, which forms a salt bridge with Asp 226 and typically places its second ring nitrogen atom within about 3 Å of a second zinc-bound oxygen of a carboxylate or phosphonate ligand. Other special circumstances can also arise. In addition, if the protein had some or all hydrogens attached, the original and prepared versions of the protein need to be compared to decide how to resolve any discrepancies.

Step four adds structural waters if any are to be kept. Nearly all waters have been removed in this work. Exceptions are 1add, 1adf, 1ebg (three waters), 1lna (two waters), and 1mdr, where a water molecule is tightly bound to a Mg^{2+} ion. By comparison, GOLD retains waters for 2ctc, 1mdr, and 1nis,[1] and FlexX does so for 1aaq, 1lna, 1xie, and 4phv.[28] Omitting structural waters can be useful because it may allow ligands to be found that are capable of displacing them. For example, we removed the water under the flaps in HIV protease to allow ligands such as the DuPont-Merck cyclic urea, which displaces this water, to dock. The resultant overly generous active site might have encouraged "false positives" in the docking. However, excellent rank orders were obtained for known HIV ligands in the database screen.[16]

The fifth step adds hydrogens to the protein, to any cofactors, and to any added structural waters, and the final step carries out a series of restrained minimizations on the protein−ligand complex. The first stage reorients repositionable side chain hydroxyls in Ser, Thr, and Tyr residues and side chain sulfhydryls of Cys. This is accomplished by tightly tethering non-hydrogen atoms (force constant = 10 kcal mol^{-1} $Å^{-2}$) and minimizing the hydrogens with torsion interactions turned off. In effect, the hydrogens are allowed to fly freely in the electrostatic wind. Subsequent steps restore the torsion potential and use progressively weaker restraints on the non-hydrogen atoms (hydrogen atoms are always free); the force constants employed are 3, 1, 0.3, and 0.1 kcal mol^{-1} $Å^{-2}$. The procedure stops when the cumulative rms deviation from the initial coordinates for non-hydrogen atoms exceeds a target value the user specifies, the default for which is 0.3 Å. The last structure having an rms deviation smaller than the target value is then selected. We perform these minimizations with either the Impact[34] or Macromodel[35] protein molecular modeling codes. When MacroModel is used, MMFF94s[25] can be employed instead of OPLS-AA.

Our experience is that this or an equivalent preparation procedure is important for attaining accurate docking with Glide. It is essential that physically untenable steric clashes often found in crystallographically determined protein sites be annealed away so that the native ligand (and others) can yield favorable vdW interaction energies for properly docked structures. It is also important that protonation states and hydrogen-bonding patterns be correct.

Acknowledgment. This work was supported in part by grants to R.A.F. from the NIH (Grants P41 RR06892 and GM 52018).

Supporting Information Available: Table S1, listing detailed results for docking ligands from 282 cocrystallized complexes into the protein sites with Glide 2.5. This ma-

terial is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Jones, G.; Wilett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a generic algorithm and an empirical binding free energy function. *J. Mol. Biol*. **1997**, *267*, 727−748.

(2) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A. A fast flexible docking method using an incremental construction algorithm. *Chem. Biol*. **1996**, *261*, 470−489.

(3) Ewing, T. J. A.; Kuntz, I. D. Critical evaluation of search algorithms for automated molecular docking and database screening. *J. Comput. Chem*. **1997**, *18*, 1175−1189. Ewing, T. J. A.; Makino, S.; Skillman, G.; Kuntz, I. D. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 411−428.

(4) Abagyan, R.; Totrov, M.; Kuznetsov, D. ICM−a new method for protein modeling and design. Application to docking and structure prediction from the distorted native conformation. *J. Comput. Chem*. **1994**, *15*, 488−506. Abagyan, R.; Totrov, M. High-throughput docking for lead generation. *Curr. Opin. Chem. Biol*. **2001**, *5*, 375−382.

(5) Welch, W.; Ruppert, J.; Jain, A. N. Hammerhead: fast, fully automated docking of ligands to protein binding sites. *Chem. Biol*. **1996**, *3*, 449−462.

(6) McMartin, C.; Bohacek, R. QXP: powerful, rapid computer algorithms for structure-based drug design. *J. Comput.-Aided Mol. Des*. **1997**, *11*, 333−344.

(7) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. Automated docking using a Lamarkian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem*. **1998**, *19*, 1639−1662.

(8) Murray, C. W.; Baxter, C. A. Frenkel, A. D. The sensitivity of the results of molecular docking to induced fit effects: Application to thrombin, thermolysin, and neuraminidase. *J. Comput.-Aided. Mol. Des*. **1999**, *13*, 547−562.

(9) McGann, M.; Almond, H.; Nicholls, A.; Grant, J. A.; Brown, F. Gaussian docking functions. *Biopolymers* **2003**, *68*, 76−90.

(10) Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. J. LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. *J. Mol. Graphics Modell*. **2003**, *21*, 289−307.

(11) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus scoring: A method of obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem*. **1999**, *42*, 5100−5109.

(12) Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem*. **2000**, *43*, 4759−4767.

(13) Stahl, M.; Rarey, M. Detailed analysis of scoring functions for virtual screening. *J. Med. Chem*. **2001**, *44*, 1035−1042.

(14) Schrödinger, L.L.C., New York. The Glide 2.5 calculations used FirstDiscovery, version 2.5021, which was released in June 2003.

(15) The results for the GOLD test set are available at http://www.ccdc.cam.ac.uk/prods/gold/rms_tab.html.

(16) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem*. **2004**, *47*, 1750−1759.

(17) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc*. **1996**, *118*, 11225−11236.

(18) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des*. **1997**, *11*, 425−445.

(19) Murphy, R. B.; Friesner, R. A.; Halgren, T. A. Unpublished research.

(20) Babine, R. E.; Bender, S. L. Molecular recognition of protein−ligand complexes: Applications to drug design. *Chem. Rev*. **1997**, *97*, 1359−1472.

(21) Whittaker, M.; Floyd, C. D.; Brown, P.; Gearing, A. J. H. Design and therapeutic application of matrix metalloproteinase inhibitors. *Chem. Rev*. **1999**, *99*, 2735−2776.

(22) Bell, I. M.; Gallicchio, S. N.; Abrams, M.; Beese, L. S.; Beshore, D. C.; Bhimnathwala, H.; Bogusky, M. J.; Buser, C. A.; Culberson, J. C.; Davide, J.; Ellis-Hutchings, M.; Fernandes, C.; Gibbs, J. B.; Graham, S. L.; Hamilton, K. A.; Hartman, G. D.; Heimbrook, D. C.; Homnick, C. F.; Huber, H. E.; Huff, J. R.; Kassahun, K.; et al. 3-Aminopyrrolidinone farnesyltransferase inhibitors: design of macrocyclic compounds with improved pharmacokinetics and excellent cell potency. *J. Med. Chem*. **2002**, *45*, 2388−2409.

(23) The reduction in the vdW radii is needed to roughly preserve hydrogen-bonding distances, which reflect a balance between attractive electrostatic and repulsive vdW forces.

(24) Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ. http://www.rcsb.org.

(25) Halgren, T. A. MMFF VI. MMFF94s option for energy minimization studies. *J. Comput. Chem*. **1999**, *20*, 720−729.

(26) For 1cps, we used the native ligand structure because the ligand contains functionality that cannot be treated by MMFF94s. The remaining four cases have larger rings that Glide's conformation generator cannot handle and for which the conformational search yielded a different ring conformation. For 1hbv and 1pro, the MMFF94s-optimized native ligand geometry was used instead, while for 1d8f and 2hct, conformational search was employed but the ring conformation was not sampled.

(27) Gasteiger, J.; Rudolph, C.; Sadowski, J. Automated generation of 3D atomic coordinates for organic molecules. *Tetrahedron Comput. Methodol*. **1990**, *3*, 537−547.

(28) Kramer, B.; Rarey, M.; Lengauer, T. Evaluation of the FlexX incremental construction algorithm for protein−ligand docking. *Proteins* **1999**, *37*, 228−241.

(29) The results for the FlexX test set are available at http://cartan.gmd.de/flexx/html/flexx-eval.html.

(30) Jain, A. N. Surflex: fully automatic flexible molecular docking using a molecular-similarity-based search engine. *J. Med. Chem*. **2003**, *46*, 499−511.

(31) CMC database is available from MDL Information Systems, Inc., San Leanrdo, CA.

(32) Oprea, T. I. Property distribution of drug-related chemical databases. *J. Comput.-Aided Mol. Des*. **1999**, *14*, 251−264. See Figure 2 in this reference.

(33) This distance is defined as the distance from a heteroatom that bears or shares a positive or negative formal charge in the ionized protein residue to the nearest ligand atom.

(34) Impact is the computational engine of Schrödinger's FirstDiscovery suite. Impact was developed in the laboratories of Prof. Ronald Levy (Rutgers University). It and the FirstDiscovery suite are available from Schrödinger, L.L.C., New York.

(35) MacroModel (formerly known as BatchMin) is a general-purpose molecular mechanics program available from Schrödinger, L.L.C., New York. MacroModel was developed in the laboratories of Prof. Clark Still (Columbia University).

JM0306430

Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening

Thomas A. Halgren,*,† Robert B. Murphy,† Richard A. Friesner,‡ Hege S. Beard,† Leah L. Frye,§ W. Thomas Pollard,† and Jay L. Banks†

Schrödinger, L.L.C., 120 W. 45th Street, New York, New York 10036, Department of Chemistry, Columbia University, New York, New York 10036, and Schrödinger, L.L.C., 1500 SW First Avenue, Portland, Oregon 97201

Received December 24, 2003

Glide's ability to identify active compounds in a database screen is characterized by applying Glide to a diverse set of nine protein receptors. In many cases, two, or even three, protein sites are employed to probe the sensitivity of the results to the site geometry. To make the database screens as realistic as possible, the screens use sets of "druglike" decoy ligands that have been selected to be representative of what we believe is likely to be found in the compound collection of a pharmaceutical or biotechnology company. Results are presented for releases 1.8, 2.0, and 2.5 of Glide. The comparisons show that average measures for both "early" and "global" enrichment for Glide 2.5 are 3 times higher than for Glide 1.8 and more than 2 times higher than for Glide 2.0 because of better results for the least well-handled screens. This improvement in enrichment stems largely from the better balance of the more widely parametrized GlideScore 2.5 function and the inclusion of terms that penalize ligand−protein interactions that violate established principles of physical chemistry, particularly as it concerns the exposure to solvent of charged protein and ligand groups. Comparisons to results for the thymidine kinase and estrogen receptors published by Rognan and co-workers (*J. Med. Chem.* **2000**, *43*, 4759−4767) show that Glide 2.5 performs better than GOLD 1.1, FlexX 1.8, or DOCK 4.01.

1. Introduction

The previous paper[1] introduced Glide,[2] a new method for rapidly docking ligands to protein sites and for estimating the binding affinities of the docked compounds. That paper described the underlying methodology and showed that Glide achieves smaller root-mean-square (rms) deviations in reproducing the positions and conformations of cocrystallized ligands than have been reported for GOLD[3] and FlexX.[4] Better docking accuracy is important in its own right in lead-optimization studies, where knowledge of the correctly docked position and conformation (pose) of a novel ligand can be crucial. In lead-discovery studies, however, docking accuracy is relevant mainly to the degree that it contributes to obtaining high enrichment in database screening. We believe that accurate scoring requires accurate docking, though accurate docking is not enough in itself.

This paper investigates the ability of Glide 2.5, run in "standard-precision" mode,[1] to identify known binders seeded into database screens for a wide variety of pharmaceutically relevant receptors. We present comparisons with earlier versions of Glide and show that very substantial progress has been made. Rigorous comparisons with other virtual screening methods are difficult for us to make because we generally do not have access either to the identical sets of decoy ligands or to other docking codes. However, we have been given access to the thymidine-kinase and estrogen-receptor datasets employed by Bissantz, Folkers, and Rognan[5] and offer comparisons to the results they published for GOLD, FlexX, and DOCK.[6]

The paper is organized as follows. In the section 2, we characterize our data sets and protocols for evaluating database enrichment. This section describes the receptors and ligands to be used, discusses certain issues concerning preparation of the receptor (most importantly, the use of reduced van der Waals radii, which is essential to achieve reasonable results in some cases), and defines the quantitative measures used to assess performance in database screening. Section 3 presents enrichment factors obtained using default parameters and describes the individual screens. Comparisons to published results for GOLD, FlexX, and DOCK for the thymidine kinase and estrogen receptors are then presented in the fourth section, and Glide's sensitivity to the choice of certain input factors is explored in section 5. Finally, the sixth section summarizes the results and discusses future directions.

2. Virtual Screening Protocol

Ligand Databases and Receptors Used. We have chosen the following nine receptors for our initial studies, five of which are represented by two or more alternative cocrystallized receptor sites:

1. thymidine kinase (1kim)
2. estrogen receptor (3ert, 1err)
3. CDK-2 kinase (1dm2, 1aq1)
4. p38 MAP kinase (1a9u, 1bl7, 1kv2)
5. HIV protease (1hpx)
6. thrombin (1dwc, 1ett)
7. thermolysin (1tmn)
8. Cox-2 (1cx2)
9. HIV-RT (1vrt, 1rt1)

* To whom correspondence should be addressed. Phone: 646-366-9555, extension 106. Fax: 646-366-9550. E-mail: halgren@schrodinger.com.
† Schrödinger, L.L.C., NY.
‡ Columbia University.
§ Schrödinger, L.L.C., OR.

10.1021/jm030644s CCC: $27.50
Published on Web 02/27/2004

The receptors for these screens cover a wide range of receptor types and therefore provide a proper test of a docking method. All were prepared using the procedure described in the preceding paper[1] or an earlier version of that procedure.

The known binders for the first two systems were specified by Rognan and co-workers,[6] while ligands for the CDK-2 kinase receptor screens and for p38 MAP kinase were provided by pharmaceutical and biotech collaborators. For thrombin, 12 of the 16 known binders were taken from the studies by Engh et al.[7] and by von der Saal et al.[8] Others are ligands for the same target protein taken from our docking-accuracy test set[1] or were developed from multitple sources in the literature.

As database ligands, we employed "druglike" decoys that averaged 400 in molecular weight (the "dl-400" dataset) in most cases. For thymidine kinase (1kim), which has a very small active site, however, we used a similar (but in this case more competitive) set with an average molecular weight of 360 (the "dl-360" dataset). The property distributions of these databases were characterized in the preceding paper.[1] We believe these compounds to be representative of the chemical sample collections of pharmaceutical and biotechnology companies. As such, they should provide a fair, and stringent, test of the efficacy of the docking method.

Each screen used 1000 database ligands and between 7 and 33 known actives. All compounds considered have 20 or fewer rotatable bonds and 100 or fewer atoms. Like the database ligands, the known binders were also MMFF94s-optimized, but in these cases, we used unbiased input geometries obtained via a MacroModel conformational search, as previously described for ligands taken from cocrystallized complexes.[1]

Glide's use of reduced atomic van der Waals radii to mimic minor readjustments of the protein (these should be distinguished from the more substantial induced-fit rearrangements modeled by the use of multiple receptor conformations) is an important issue in the setup of the docking runs. Glide currently supports uniform van der Waals scaling of the radii of nonpolar protein and/or ligand atoms. To characterize the performance that can be expected when Glide is run "out of the box", the principal results presented in this paper use default 1.0 protein scaling (which means that the OPLS-AA van der Waal (vdW) radii are not changed) and 0.8 ligand scaling; the same scalings were used to assess docking accuracy.[1] In the fifth section, we compare these results with results obtained using scaling factors identified in earlier docking studies with Glide as giving optimal results. The comparisons show that default scaling performs well, though optimizing the scaling factors can improve the performance in some cases.

Measures of Virtual Screening Effectiveness. To quantify Glide's ability to assign high ranks to ligands with known binding affinity, we report enrichment factors in graphical and tabular form and present accumulation curves that show how the fraction of actives recovered varies with the percent of the database screened. Following Pearlman and Charifson,[9] the enrichment factor can be written as

$$\mathrm{EF} = \frac{\mathrm{Hits}_{\mathrm{sampled}}/N_{\mathrm{sampled}}}{\mathrm{Hits}_{\mathrm{total}}/N_{\mathrm{total}}} \quad (1)$$

Equivalently, this can be written as

$$\mathrm{EF} = \{N_{\mathrm{total}}/N_{\mathrm{sampled}}\}\{\mathrm{Hits}_{\mathrm{sampled}}/\mathrm{Hits}_{\mathrm{total}}\} \quad (2)$$

Thus, if only 10% of the scored and ranked database (i.e., $N_{\mathrm{total}}/N_{\mathrm{sampled}} = 10$) needs to be assayed to recover all of the $\mathrm{Hits}_{\mathrm{total}}$ actives, the enrichment factor would be 10. But if only half of the total number of known actives are found in this first 10% (i.e., if $\mathrm{Hits}_{\mathrm{sampled}}/\mathrm{Hits}_{\mathrm{total}} = 0.5$), the effective enrichment factor would be 5.

Equation 2 is useful when sampling an initial, small fraction of a database, but to measure performance for recovering a substantial fraction of the active ligands, we prefer to modify the definition of enrichment as follows:

$$\mathrm{EF}' = \{50\%/\mathrm{APR}_{\mathrm{sampled}}\}\{\mathrm{Hits}_{\mathrm{sampled}}/\mathrm{Hits}_{\mathrm{total}}\} \quad (3)$$

In this equation, $\mathrm{APR}_{\mathrm{sampled}}$ is the average percentile rank of the $\mathrm{Hits}_{\mathrm{sampled}}$ known actives. Intuitively, this makes sense: if the actives are uniformly distributed over the entire ranked database, the average percentile rank for an active would be 50% and the enrichment factor would be 1. Unlike eqs 1 and 2, however, this formula considers the rank of each of the $\mathrm{Hits}_{\mathrm{sampled}}$ known actives, not just the rank of the last active found (which is what N_{sampled} is likely to be). As a result, the enrichment factor will be larger than the value computed from eq 1 or 2 if the actives are concentrated toward the beginning of the N_{sampled} ranked positions but will smaller if the actives are grouped toward the end of this list. This is appropriate because a key objective in database screening is to find active compounds as early as possible in the ranked database; the new definition is better at indicating when this is happening.

3. Virtual Screening Results

Overview of Glide's Performance in Database Screening. Table 1 compares the performance of Glide 1.8, 2.0, and 2.5 SP (standard-precision)[1] using as definitions of enrichment EF′(70), which measures the enrichment for recovering 70% of the known actives, and EF(2%), which measures enrichment for assaying the top 2% of the ranked database. These represent "global" and "early" enrichment, respectively. Though 70% recovery is arbitrary, we feel this is a realistic standard for docking into a rigid protein site, given that such a site is unlikely to be properly shaped to house all the known actives when the site is relatively plastic. For use in virtual screening, it can be crucial to concentrate as many active ligands as possible in the topmost portion of the ranked database. Because the present screens use only ~1000 ligands, however, 2% (20 ranked positions or so) is about the smallest percentage we can examine, given that we have roughly 10−30 known actives to place. We also report enrichment factors EF(5%) and EF(10%) for Glide 2.5. Note that the maximum attainable enrichment factors are 50, 20, and 10, respectively, for EF(2%), EF(5%), and EF(10%). Also listed are average enrichment factors computed using a generalized geometric mean that weights the smaller enrichment factors more heavily.[10] (For example, the geometric mean of 1 and 25 is 5, not 13.) This weighting

Table 1. Comparison of Enrichment Factors for Glide 1.8, Glide 2.0, and Glide 2.5[a]

		EF′ (eq 3) 70% recovery			EF (eq 2) 2% of database			EF (eq 2) 5%	EF (eq 2) 10%
screen	site	GS 1.8	GS 2.0	GS 2.5	GS 1.8	GS 2.0	GS 2.5	GS 2.5	GS 2.5
thymidine kinase (tk)	1kim	4.2	7.6	17.2	0.0	10.0	25.0	12.0	9.0
tk-pyrimidine ligands	1kim	4.5	6.7	20.4	0.0	7.1	28.6	14.3	8.6
estrogen receptor	3ert	88.4	79.8	66.9	35.0	35.0	35.0	14.0	8.0
estrogen receptor	1err	88.4	37.5	46.7	35.0	30.0	30.0	14.0	9.0
CDK-2 kinase	1dm2	3.6	3.9	6.8	5.0	10.0	15.0	10.0	6.0
CDK-2 kinase	1aq1	2.1	3.8	5.4	5.0	15.0	10.0	8.0	6.0
p38 MAP kinase	1a9u	2.0	1.8	2.5	0.0	2.9	5.9	4.7	2.4
p38 MAP kinase	1bl7	1.8	2.9	3.5	2.9	5.9	8.8	3.5	4.1
p38 MAP kinase	1kv2	4.5	2.9	4.8	8.8	11.8	8.8	8.2	4.1
HIV protease	1hpx	10.8	7.8	47.6	30.0	13.3	36.7	17.3	8.7
thrombin	1dwc	5.8	2.8	12.1	6.2	3.1	15.6	11.2	8.1
thrombin	1ett	5.2	5.6	38.0	12.5	3.1	34.4	16.2	9.4
thermolysin	1tmn	1.6	15.2	24.5	5.0	15.0	25.0	18.0	9.0
Cox-2	1cx2	3.6[b]	3.4[b]	5.0[b]	7.6	3.0	13.6	7.9	5.2
Cox-2 (site-1 ligands)	1cx2	5.7	5.7	13.9	10.9	4.3	19.6	11.3	7.4
HIV rev. transcriptase	1rt1	4.6	3.2	8.2	3.0	0.0	12.1	10.3	6.4
HIV-RT	1vrt	1.8	2.0	7.1	0.0	0.0	7.6	9.1	5.8
av enrichment factor:		5.3	6.5	14.8	6.0	7.1	18.6	11.4	7.0

[a] Enrichment factors can be at most 50 for 2% sampling, 20 for 5% sampling, and 10 for 10% sampling. [b] EF′(60) value.

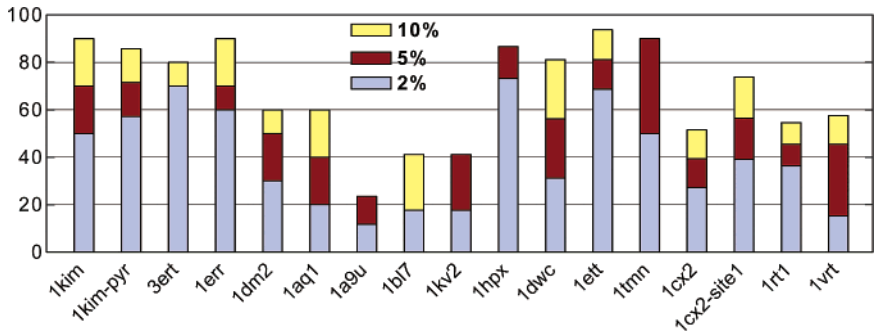

Figure 1. Percent of actives recovered with Glide 2.5 for assaying 2%, 5%, and 10% of the ranked database for the screens considered in this paper. The PDB codes are defined in Table 1.

is appropriate because what is most needed is a method that can be counted on to always perform reasonably well rather than one that does very well for some systems but is useless for others.

Table 1 shows that Glide 2.5 is much better than its predecessors at identifying active ligands. Because of improvements to the more difficult screens, both global and early enrichment have tripled since the release of Glide 1.8. The CDK-2 and p38 screens are still problematic, but thymidine kinase now does well, and thrombin, HIV protease, thermolysin, Cox-2, and HIV-RT all have improved substantially.

Figure 1 summarizes Glide's ability to rank active ligands in the first 2%, 5%, and 10% of the scored and ranked database. In many, though not all, cases, a significant fraction of the actives are found in the top 2% of the database.

Finally, Figure 2 displays the percent of known actives recovered as a function of the percent of the ranked database sampled for Glide 1.8, 2.0, and 2.5. This complementary view also shows that Glide 2.5 performs exceedingly well for many of the targets and also highlights cases in which further improvement is particularly desirable.

Detailed Results for Database Screens. The remainder of this section describes the individual database screens and presents graphical depictions of enrichment for Glide 1.8, 2.0, and 2.5. Detailed listings of the ranks of the active ligands, their GlideScore values, their hydrogen-bonding scores, and their Coulomb−vdW interaction energies with the protein site are available in Supporting Information (Tables S1−S15).

1. Thymidine Kinase (1kim). Rognan and co-workers studied the binding of 10 known thymidine kinase ligands to the protein from the 1kim complex.[6] For database ligands, they used 990 randomly chosen compounds from a filtered version of the ACD database. Only the 1kim ligand (dT) and one other ligand (idu) are reported to be submicromolar binders. The others (five are also pyrimidine derivatives and three are purines (acv, gcv, pcv)) range in activity from 1.5 to 200 μM. Realistically speaking, computational screening of compound databases usually can only hope to discover micromolar ligands. This poses a stiff challenge because Charifson et al.[11] found that the docking methods they surveyed performed reasonably well at finding low-nanomolar binders seeded into a database screen but fell off rapidly in efficacy as the activity of the known binders decreased. The ability to identify micromolar binders in a database screen is therefore a stringent and relevant test.

Figures 2a and 3 examine the thymidine kinase screen. In this case, Figure 3a uses the seven pyrimidine and three purine-based ligands defined by Rognan and co-workers as known actives while Figure 3b uses only the seven pyrimidines. Comparison of the lowest bar segments shows that Glide 2.5 is significantly more effective than Glide 1.8 or 2.0 at concentrating known

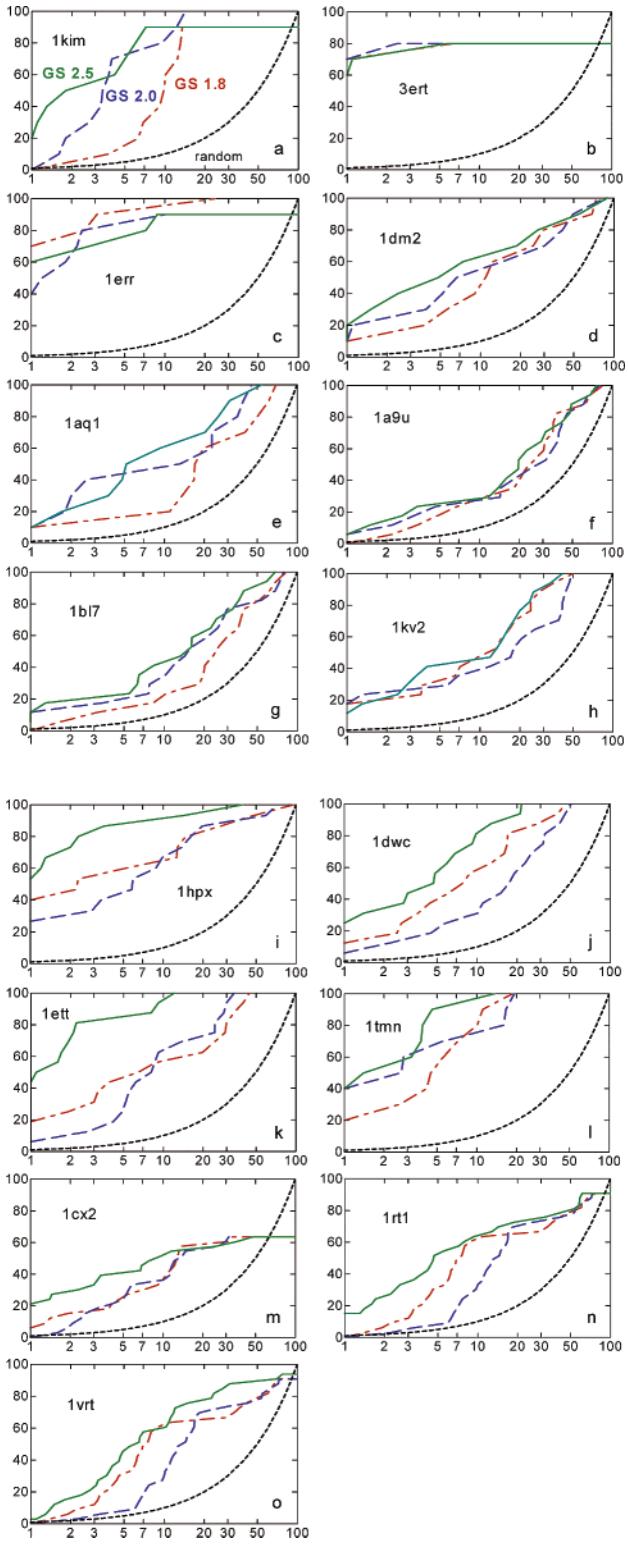

Figure 2. Percent of known actives found (*y* axis) vs percent of the ranked database screened (*x* axis) for Glide 2.5 (solid green), Glide 2.0 (blue dashed), and Glide 1.8 (red dot−dashed). Black dotted lines show results expected by chance. The listed PDB codes are defined in Table 1.

actives in the first 2% of the ranked database for both the mixed and pyrimidine-only screens; better performance relative to earlier verisons of Glide is also shown in Figure 2a for the mixed screen.

The reason for separating out the pyrimidines is that a labile Gln 125 side chain undergoes a 180° rotation

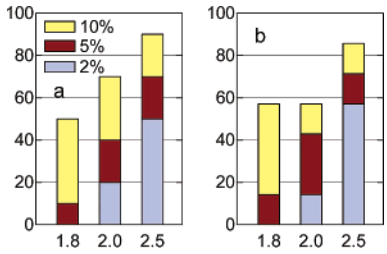

Figure 3. (a) Percent of thymidine kinase (1kim) actives recovered with Glide 1.8, 2.0, and 2.5 for assaying the first 2%, 5%, and 10% of the ranked database. (b) Percent recovered using only the seven pyrimidine-based ligands as actives.

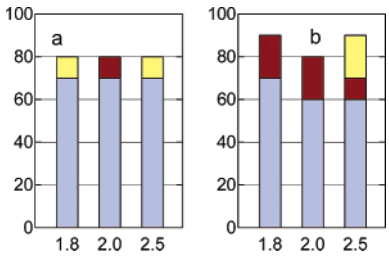

Figure 4. Percent of estrogen receptor actives recovered for assaying the top 2%, 5%, and 10% of the ranked database: (a) 3ert site; (b) 1err site.

on going from 1kim, a pyrimidine site, to one of the purine sites (2ki5, 1ki2, 1ki3). This rotation essentially exchanges the terminal NH_2 and C═O groups and means that purines cannot dock properly into a pyrimidine site, nor can pyrimidines dock properly into a purine site; in each case, the geometry that is correct for the parent site has an acceptor−acceptor and/or a donor−donor clash in the alternative site. This incompatibility results in larger rms deviations for cross-docking, as reported by Rognan and co-workers,[6] by Jain,[12] and by us.[1] Nevertheless, the misdocked purines find many favorable interactions and score well (Table S1, Supporting Information) because the terms in the standard-precision version of GlideScore 2.5 that penalize breaches of complementarity are too small to significantly penalize the misdocked purine structures. In contrast, Extra-Precision docking[13] imposes larger penalties for docking mismatches and therefore is more likely to be capable of rejecting ligands that do not dock properly into the site.

2. Estrogen Receptor (3ert, 1err). The target proteins for the estrogen receptor (ER) screen are the 3ert receptor site studied by Rognan and co-workers[6] and the 1err site used by Stahl and Rarey;[14] the native ligands are 4-hydroxytamoxifen and raloxifene, respectively. Both sites are open enough to dock antagonists as well as agonists. Our studies used the 10 low-nanomolar ERα antagonists that Rognan selected as active binders. This is one case in which the nonbonded radii need to be scaled down to allow the known binders to dock correctly. For example, five of the known binders had positive Coulomb−vdW interaction energies when no scaling was done. For Glide 1.8 and 2.0, we originally used 0.9 protein/0.8 ligand scaling, but we employ the default 1.0/0.8 scaling here.

Figures 2b,c and 4 show that both estrogen-receptor sites are treated very well by all three scoring functions. Tables S2 and S3 (Supporting Information) list the Glide 2.5 rankings.

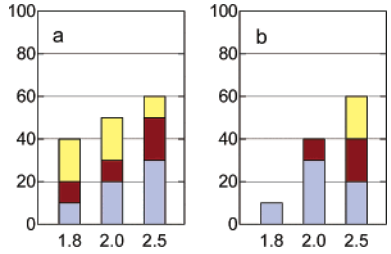

Figure 5. Percent of CDK-2 kinase actives recovered for assaying the top 2%, 5%, and 10% of the ranked database: (a) 1dm2 site; (b) 1aq1 site.

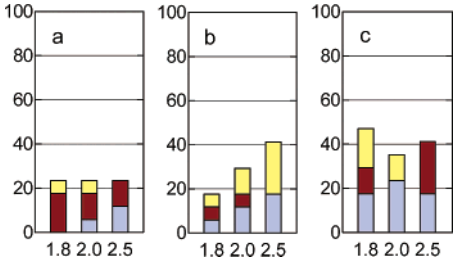

Figure 6. Percent of p38 MAP kinase actives recovered for assaying the top 2%, 5%, and 10% of the ranked database: (a) 1a9u site; (b) 1bl7 site; (c) 1kv2 site.

3. CDK-2 Kinase (1dm2, 1aq1). The CDK-2 kinase site is highly flexible. A key issue is the length of the rather narrow binding cavity, which if insufficient will prevent many active ligands from correctly docking. After examining superimposed structures for five co-crystallized PDB complexes, we chose a site from the 1dm2 complex that is more elongated than most, though not the most generous. To investigate the sensitivity of the docking to the choice of receptor site, we chose the slightly more open 1aq1 receptor as a second site. In each case, we used default 1.0/0.8 scaling.

Figures 2d,e and 5 show that Glide 2.5 outperforms its predecessors for the 1dm2 site and greatly outperforms them for 1aq1. The Glide 2.5 rankings and scorings are given in Tables S4 and S5 (Supporting Information).

4. p38 MAP kinase (1a9u, 1bl7, 1kv2). The p38 active site is particularly prone to alter its shape upon ligand binding. Therefore, we studied three different PDB receptor structures: 1a9u, 1bl7, and 1kv2. The 1kv2 site exhibits a particularly large alteration of the ligand-free structure in that a long loop undergoes a substantial change in conformation when its native ligand binds.[15] This screen employs 14 known p38 binders supplied by a colleague from the biotechnology industry in addition to the cocrystallized ligands from the 1a9u, 1bl7, and 1kv2 structures.

The plasticity of the p38 active site makes it difficult to dock a large number of active compounds properly into any single version of the receptor structure and hence leads to relatively small values of global enrichment such as EF′(70). Figures 2f−h and 6 show that the 1kv2 site is the most amenable one for the particular selection of active compounds used in this study. Glide 2.5 achieves relatively little improvement over previous versions in this case, in part because the p38 site is large and requires a very hydrophobic binding mode (in general, only one to two hydrogen bonds are made by correctly docked p38 actives). Such sites represent a severe challenge for most empirical scoring functions.

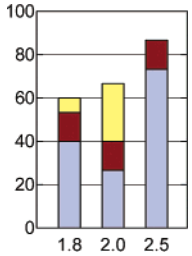

Figure 7. Percent of actives recovered for assaying the top 2%, 5%, and 10% of the ranked database for the 1hpx site of HIV protease.

We can report, however, that we have made significant progress in handling sites of this nature in the ongoing development of Extra Precision Glide.[13]

Detailed results for Glide 2.5 are shown in Tables S6−S8 (Supporting Information).

5. HIV Protease (1hpx). Our screening database contains 15 ligands from cocrystallized HIV-1 protease complexes included in our docking-accuracy test set.[1] Here, we focus on 1hpx as the target. The 1hpx complex retains the usual water under the "flaps" that enclose the active site, but we removed it so that ligands that displace this water, such as XK263 (from the 1hvr complex) and A-98881 (from 1pro), could dock. The removal of this water raises the question of whether the resultant overly generous site might recognize a large number of false positives. However, this proved not to be the case.

As with the estrogen-receptor screen, our original docking experiments used 0.9 protein/0.8 ligand scaling. However, this site is not especially "tight" and default 1.0/0.8 scaling works quite well, especially for Glide 2.5 (cf. Figures 2i and 7). Indeed, 7 ligands are found in the top 10 ranked positions and 12 are found in the top 20 (Table S9, Supporting Information). This is good performance by any standard.

6. Thrombin (1dwc, 1ett). Our docking-accuracy test set[1] contains five unique thrombin inhibitors (1dwc, 1etr, 1dwb, 1dwd = 1ets = 1ppc, and 1ett). However, only four are highly active because the 1dwb ligand, benzamidine, is too small to bind tightly (the experimental binding affinity is −5.4 kcal/mol [1]). To supplement these four binders, we included the 12 thrombin inhibitors from Engh et al.[7] and von der Saal et al.[8] that have reported binding affinities of 10 μM or better, bringing the total number of actives to 16.

To prepare for the original 1dwc screen for Glide 1.8, we found that all combinations of 0.8, 0.9, and 1.0 vdW scaling for nonpolar protein and ligand atoms gave strongly negative Coulomb−vdW energies for the known actives and afforded reasonably negative GlideScores. However, the model that used unscaled vdW radii for both the protein and the ligand gave the best overall GlideScores and yielded an unfavorable hydrogen-bonding score for only one ligand. We therefore chose to not scale the vdW radii. The present results, however, use default 1.0 protein/0.8 ligand scaling. Thus, this screen differs in the opposite sense from the estrogen-receptor and HIV-protease screens, which originally used greater scaling than the current default. As we show in the section 5, the two scaling models yield comparable results. Thus, employing larger scaling than is needed to allow the known actives to fit into the site

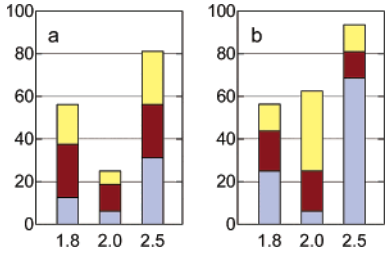

Figure 8. Percent of thrombin actives recovered for assaying the top 2%, 5%, and 10% of the ranked database: (a) human thrombin, 1dwc site; (b) bovine thrombin, 1ett site.

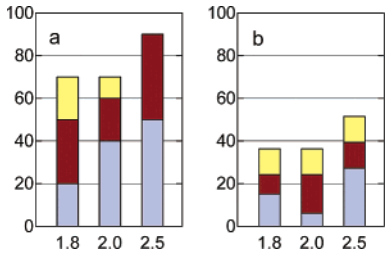

Figure 9. Percent of actives recovered for assaying the top 2%, 5%, and 10% of the ranked database: (a) thermolysin, 1tmn site; (b) Cox-2, 1cx2 site.

did not degrade Glide's ability to rank the known binders highly in this case.

Figures 2j,k and 8 show that Glide 2.5 performs better than either Glide 1.8 or 2.0. The same qualitative trend is seen in the EF′(70) and EF(2%) enrichment factors reported in Table 1. The comparisons show that the 1ett site consistently yields better results. Tables S10 and S11 (Supporting Information) list the Glide 2.5 rankings.

7. Thermolysin (1tmn). As target receptor, we chose the protein from the 1tmn complex. The 1tmn ligand, known as CLT for "carboxy-leu-trp", coordinates with the active-site Zn^{2+} ion via a carboxylate group and places its leucine side chain into a hydrophobic pocket. Including 1tmn, the docking-accuracy test set contains 13 thermolysin complexes, including one, 1lna, in which Co^{2+} replaces Zn^{2+}. However, the 2tmn, 4tln, and 1hyt ligands have only about 25 atoms (including hydrogens) and are too small to bind tightly; for example, the measured binding affinities for 2tmn and 4tln are only −5 to −7 kcal/mol.[1] In this study, we use the 10 larger, drug-sized ligands. Thermolysin is known to have a rigid active site, so the choice of target structure is probably unimportant in this case.

As for thrombin, preliminary studies suggested a preference for using unscaled radii for both the protein and the ligand (i.e., 1.0/1.0 scaling), but the results presented here use default 1.0/0.8 scaling. That unscaled radii yield good results seems clearly related to the rigid and open nature of the site. In this case, somewhat better results are obtained with 1.0/1.0 scaling (see section 5), but default scaling also does well. As Figures 2l and 9a show, Glide 2.5 performs much better than Glide 1.8 and somewhat better than Glide 2.0. For Glide 2.5, 5 of the 15 top-ranked ligands are known binders (Table S12, Supporting Information).

One key to the improved performance is that GlideScore 2.5 specifically rewards metal ligation by anionic ligand functionality. We made this change on the basis of experimental evidence that metalloproteases strongly favor anionic ligands.[16,17] A second element is that GlideScore 2.5 considers only the single strongest interaction when the ligation is bi- or multidentate. A third is that Glide 2.5 reduces the net ionic charges for most charged−charged and charged−polar interactions but leaves them unchanged for metal−ligand interactions. Thus, the Coulombic contribution to GlideScore 2.5 uses the full Zn^{2+}−ligand interaction energy, further helping to differentiate anionic ligands.

These elements were also included in GlideScore 2.0, but GlideScore 2.5 goes one step further by recognizing that neutral ligands such as imidazoles can be effective binders when Zn^{2+} and the "tripod" of protein residues on which it sits are net neutral (e.g., when Zn^{2+} is coordinated by two glutamates and a histidine, as in farnesyl protein transferase,[18] rather than by one glutamate and two histidines, as in thermolysin). In such cases, the term in GlideScore 2.5 that rewards a geometrically appropriate metal ligation by −2.0 kcal/mol when the ligand functionality is anionic is omitted when the apo site is net-neutral. While this modification seems appropriate, further studies will be needed to determine whether it gets the "balance" right for net-neutral sites.

8. Cox-2 (1cx2). We obtained structures for 33 known binders from the literature. These include the native 1cx2 ligand (SC-558, ligand 24), celecoxib (ligand 2), rofecoxib (ligand 3), indomethacin (ligand 10), deprotonated indomethacin (ligand 26), flurbiprofen (ligand 25), deprotonated flurbiprofen (ligand 33), ML-3000 (ligand 11), and deprotonated ML-3000 (ligand 31). Examination of various combinations of protein and ligand scaling factors using Glide 1.8 led us to select 1.0 protein/0.8 ligand scaling, which we also used here and for Glide 2.0.

These Glide dockings have one unusual feature, namely, that only 23 of the 33 known actives dock with negative Coulomb−vdW energies, even with relatively heavy scaling of protein and ligand nonpolar vdW radii, when a "normally sized" docking box centered around the 1cx2 ligand is used. When the docking box is made much larger, the remaining 10 ligands dock "successfully" but occupy a site that is displaced by 10−12 Å from the primary site. However, the "site 2" ligands have relatively poor Coulomb−vdW interaction energies and often make no hydrogen bonds. It thus seems unlikely that these Cox-2 ligands actually dock into this second site. The most likely explanation is that some variable element in the site geometry is responsible and that the limitations of docking to a rigid site are particularly extreme in this case.

Figures 2m and 9b show that Glide 2.5 performs much better than Glide 1.8 or 2.0. It does particularly well for early enrichment, as indicated by the lowest segments of the bar chart in the latter figure. Indeed, Glide 2.5 places 9 of the actives in the first 20 positions in the ranked database (Table 13S, Supporting Information). Figure 9b and Table 1 show that the calculated enrichment factors are relatively low when based on all 33 Cox-2 ligands. When recomputed to count only the 23 "site 1" ligands as actives, however, Glide 2.5 yields a quite decent EF′(70) value of 14.2. Thus, Glide 2.5 is effective at finding active Cox-2 ligands; it just cannot

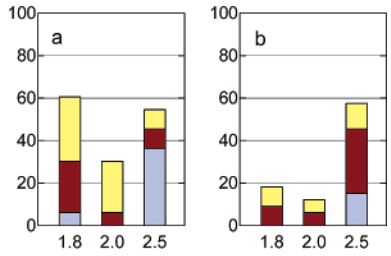

Figure 10. Percent of HIV reverse transcriptase actives recovered for assaying the top 2%, 5%, and 10% of the ranked database: (a) 1rt1 site; (b) 1vrt site.

dock and score all of them well when the 1cx2 site is used.

9. HIV-RT (1vrt, 1rt1). For HIV reverse transcriptase, we docked a set of 33 active ligands into the non-nucleoside binding site used by nevirapine, Sustiva, and other NNRTI compounds. The active ligands, taken from a variety of literature sources, include nevirapine and MKC-442. We used both the nevirapine site (1vrt) and the MKC-442 site (1rt1) as targets. On the basis of the dockings of the active ligands, we chose 0.9 protein/0.8 ligand scaling for 1rt1 and 1.0 protein/0.8 ligand scaling for 1vrt when testing Glide 2.0. The present results, however, use default 1.0/0.8 scaling for both sites.

Table 1 and Figures 2n,o and 10 show that Glide 2.5 greatly outperforms the earlier releases for these two HIV-RT sites. This site is also very hydrophobic and offers few hydrogen-bonding opportunities (cf. the hydrogen-bonding scores in Tables S14 and S15, Supporting Information). It therefore qualifies as a difficult site for an empirical scoring function such as GlideScore 2.5. In such a site it is crucial to recognize and penalize mismatches in complementarity in the docked poses. The improved performance relative to the earlier versions of Glide reflects the better balance of the more widely parametrized GlideScore 2.5 function as well as the inclusion of desolvation penalties; these terms play an even larger role in the Extra-Precision Glide 2.5 scoring function.[13]

4. Comparison to Other Methods

Comparisons usually are difficult for us to make because we do not have access to other docking codes and because published comparisons often use proprietary datasets.[11,14,19] In this section, however, we present comparisons to published results for GOLD 1.1,[3] FlexX 1.8,[4] and DOCK 4.01[20,21] for the thymidine kinase and estrogen receptors[6] using datasets provided to us by D. Rognan.[5] We caution that the earlier versions of GOLD, FlexX, and DOCK used by Rognan and co-workers may not be representative of the current capabilities of these methods. However, the same comparisons were also used by Jain in his recent paper introducing the Surflex method.[12]

Their study used GOLD, FlexX, and DOCK as docking engines and employed ChemScore,[22] FlexX, the DOCK energy score,[23] GOLD, PMF,[24] Fresno,[25] and Score[26] to rank the docked poses. The best result, obtained by a few of the 21 combinations of 3 docking engines and 7 scoring functions, found 8 of 10 actives in the first 8−10% of the ranked database. Given that only 0.8−1.0 actives would be found by chance, this performance corresponds to an enrichment factor of roughly 10. The best single models were DOCK with PMF scoring, FlexX with PMF scoring, and GOLD with GOLD scoring. Many models performed poorly, however. For example, when GOLD was used as the docking engine, ChemScore, the DOCK energy score, and Fresno all found just one active in the first 60% of the database, whereas six would be found by chance. Rognan and co-workers also found that some of the docking-method/scoring-functions combinations did poorly for the estrogen receptor, though others did well.

Table 2. Comparison of Enrichment Factors EF′(70)[a] for Glide, GOLD, FlexX, and DOCK for the Thymidine Kinase Receptor (1kim) and the Estrogen Receptor (3ert), Using Data Sets of Rognan and Co-workers[5,6]

	EF′(70), 70% recovery of known actives					
screen	Glide 1.8	Glide 2.0	Glide 2.5	GOLD[b] 1.1	FlexX[b] 1.8	DOCK[b] 4.01
thymidine kinase (1kim)	4.2	11.7	19.3	8.2	11.1	3.0
estrogen receptor (3ert)	70.0	72.1	47.1	28.5	8.9	6.7
av (geometric mean)	17.1	29.0	30.2	15.3	9.9	4.5

[a] Equation 3. [b] Reference 6.

Our calculations used Rognan's receptor and ligand preparations and employed the 0.9 protein/0.8 ligand scaling of nonpolar vdW radii we recommend for use when the protein site has not been relaxed to remove possible steric clashes (see section 5). We also used docking and scoring grids of the same size employed in section 3. As previously noted, Rognan and co-workers randomly selected the 990 database ligands from a filtered version of the ACD database. They employed various scoring functions in conjunction with each of the docking methods, but we focus here on the native GOLD-docking/GOLD-scoring, FlexX-docking/FlexX scoring, and DOCK-docking/DOCK-scoring combinations. These are the ones most likely to be used in a pharmaceutical setting, where project needs may not permit extensive explorations of alternative docking/scoring combinations to be carried out.

Comparisons of docking results for Glide, GOLD, FlexX, and DOCK are presented in Table 2 and in Figures 11 and 12. These comparisons show that DOCK docking followed by DOCK-energy scoring is the worst model. Glide 2.5 appears to be the best model overall by virtue of its superior performance for thymidine kinase, though Surflex does even better for this receptor[12] and though Glide 1.8 and 2.0 do better for the estrogen receptor (cf. Figure 12 and Table 2). The latter may reflect the better balance of the Glide 2.5 scoring function, which often yields better results for poorly treated screens at the cost of some degradation in performance for well-handled screens. FlexX and GOLD show decent enrichment, but neither is as effective as Glide.

5. Sensitivity to vdW Scaling Parameters

As noted previously, Glide by default leaves the protein radii unchanged but scales the nonpolar ligand radii by 0.8; we refer to this as 1.0/0.8 scaling. Scale factors smaller than 1.0 make the protein site "roomier"

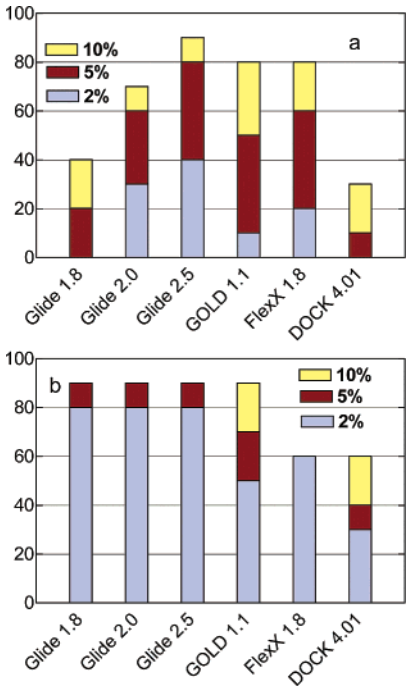

Figure 11. Percent of actives recovered by Glide 1.8, 2.0, and 2.5 and by GOLD, FlexX, and DOCK for assaying the first 2%, 5%, and 10% of the ranked database, using the datasets of Rognan and co-workers:[5,6] (a) thymidine kinase (1kim); (b) estrogen receptor (3ert).

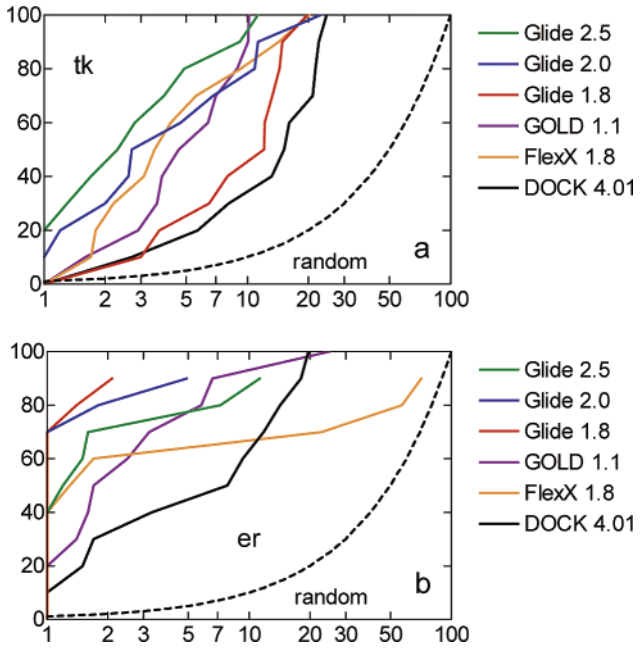

Figure 12. Percent of known actives found (*y* axis) vs percent of the ranked database screened (*x* axis) for Glide 2.5 (green), Glide 2.0 (blue), and Glide 1.8 (red) and for GOLD (purple), FlexX (orange), and DOCK (black) using the datasets of Rognan and co-workers:[5,6] (a) thymidine kinase (1kim); (b) estrogen receptor (3ert).

by "moving back" the surface of nonpolar regions of the protein and/or ligand. These adjustments emulate to some extent the effect of "breathing" motions a protein site might make to accommodate a tight-binding ligand that is slightly larger than the native, cocrystallized ligand. Cross-docking tests have consistently shown that it is important to modify the final vdW surface in this manner. Too much scaling, however, is detrimental because active ligands may no longer make suitably specific interactions with the receptor if the cavity is too large. Moreover, ligands that are too large to bind to the physical receptor may begin to dock and score well computationally, swelling the ranks of the false positives. The object is to find a "happy medium" between too little scaling and too much.

Table 3 shows how the choice of scaling factors effects the enrichment obtained in the database screens described in section 3. Thermolysin is an example of a rigid, open site, while the p38 MAP kinase site is highly mobile and the estrogen receptor contains a tightly enclosed hydrophobic channel. Alternative scalings are shown for cases in which the preferred scaling previously found for Glide 1.8 or 2.0 differs from the current default scaling. The table shows that the new default scaling works as well as or better than the previously identified preferential scaling in six of the eight cases. The original scaling gives substantially better enrichment factors for 1err and performs somewhat better for 1tmn, but the enrichment factors are high in these cases and the default enrichments are also good.

Table 4 shows that 0.9/0.8 scaling occasionally allows one or two additional actives to dock. Moreover, the more generous scaling almost always produces a significantly lower rank for the last "common" active found (e.g., the 8th for 3ert, the 9th for 1err, or the 21st for 1cx2). This, too, indicates that 0.9/0.8 scaling produces a better physical model when the fit is tight. The 1rt1 screen is an exception because 0.9/0.8 is not in fact the optimal scaling for Glide 2.5.

The conclusion we draw is that use of optimal scaling factors should be considered for "high-performance" screens. When active ligands are unavailable or will not be used to determine the scaling factors, the current default should normally be used. However, if the protein heavy atom coordinates are taken directly from the X-ray structure, it may be better to use 0.9/0.8 scaling to reduce the effect of unresolved steric clashes. This more generous scaling should also be used in cases in which it is known that the active-site region is tight and enclosed (an example being the hydrophobic channel of the estrogen receptor) because it will be difficult in such cases for certain active ligands to avoid serious steric clashes with the rigid site. Conversely, a lesser degree of scaling might be tried if the site is open and is known to be relatively rigid.

6. Discussion and Conclusions

This paper has presented results for 15 database screens covering 9 widely varying receptor types. Using recovery of 70% of the known actives as a benchmark, Glide 2.5 yields enrichment factors of at least 10 for all but CDK-2, p38, Cox-2, and HIV-RT and of less than 5 for only the 1a9u, 1bl7, and 1kv2 sites for p38. For Cox-2, the modest EF′(60) value found when all 33 actives are considered is mitigated by the finding that Glide 2.5 places 9 of the 33 known binders in the first 20 ranked positions. HIV-RT has long been problematic, but Glide 2.5 treats it considerably better than did its predecessors. Two of the most troublesome remaining screens are p38 and CDK-2, but progress is observed here, too, when XP Glide is employed.[13]

Table 3. Sensitivity of Calculated Enrichment Factors to vdW Scaling Sactors[a]

		vdW scaling		enrichment factor			
screen	site	protein	ligand	EF(2%)	EF(5%)	EF(10%)	EF′(70)
thymidine kinase (tk)	1kim	1.0	0.9	20.0	12.0	9.0	17.9
		1.0	0.8	25.0	12.0	9.0	17.9
tk-pyrimidine[a]		1.0	0.9	21.4	14.3	8.6	21.9
		1.0	0.8	28.6	14.3	8.6	22.5
estrogen receptor	3ert	0.9	0.8	40.0	16.0	8.0	70.7
		1.0	0.8	35.0	14.0	8.0	75.0
estrogen receptor	1err	0.9	0.8	35.0	18.0	9.0	60.4
		1.0	0.8	30.0	14.0	9.0	41.2
thrombin	1dwc	1.0	1.0	12.5	10.0	6.9	10.3
		1.0	0.8	15.6	11.2	7.5	11.6
HIV protease	1hpx	0.9	0.8	36.7	18.7	9.3	40.1
		1.0	0.8	40.0	17.3	8.7	46.0
thermolysin	1tmn	1.0	1.0	25.5	20.0	10.0	30.5
		1.0	0.8	25.0	18.0	9.0	24.5
HIV-RT	1rt1	0.9	0.8	6.1	7.9	6.4	6.4
		1.0	0.8	13.6	10.9	6.4	9.1

[a] In each case, the preferential scaling model used with Glide 2.0 is listed first.

Table 4. Number of Known Actives Docked with Negative Coulomb−vdW Interaction Energies as a Function of the Protein and Ligand vdW Scale Factors for Nonpolar Atoms

			no. docked		rank of last common active	
screen	site	no. of actives	1.0/0.8	0.9/0.8	1.0/0.8	0.9/0.8
estrogen receptor	3ert	10	8	9	58	17
estrogen receptor	1err	10	9	9	87	43
HIV protease	1hpx	15	15	15	408	204
Cox-2	1cx2	33	21	23	490	355
HIV-RT	1rt1	33	30	30	632	735

Comparison to results obtained for Glide 1.8 and 2.0 shows that average measures for both early and global enrichment are 2−3 times higher for Glide 2.5. Most importantly, Glide 2.5 performs significantly better for many of the more difficult screens; this qualitative improvement should be borne in mind when assessing comparitive studies based on Glide 1.8 or 2.0, which have begun to appear.[19] The improved enrichment stems partly from the inclusion of scoring-function terms that penalize ligand−protein interactions that violate established principles of physical chemistry, particularly as it concerns the exposure to solvent of charged protein and ligand groups. Given reports we have received from users that earlier versions of Glide were at least competitive in database enrichment to other commercially available methods, these results suggest that Glide 2.5 may represent a qualitative advance in scoring accuracy and virtual screening efficiency. Comparisons made to GOLD 1.1, FlexX 1.8, and DOCK 4.01 for the thymidine kinase and estrogen receptors using datasets prepared by Rognan and coworkers support this view, though we again caution that these comparisons may not be representative of the current capabilities of these methods.

Glide 2.5 has a number of advantages relative to previous versions. One is that generally good results are obtained with the new default 1.0 protein/0.8 ligand scaling. Calibrating the scale factors can lead to improved performance, but this may be less critical than with earlier versions of Glide, which employed less well-balanced scoring functions. A second advantage is that hydrogen-bond filters (i.e., imposition of a cutoff on the hydrogen-bond energy) and/or metal-ligation filters are no longer necessary. These elements broaden the range of applicability of Glide and simplify its use.

One theme that runs consistently through the results is that Glide does best when the active ligands make multiple hydrogen bonds to the receptor and does worst when the site is hydrophobic and offers few such opportunities. From what we have seen in the literature, this behavior is not unique to Glide. One of the key problems in database screening (one on which we have made considerable progress in ongoing work with XP Glide[13]) is how to properly model binding when it is mainly hydrophobic in character. These new developments will be described in a subsequent paper.

The most challenging problem in the use of docking methods in pharmaceutical applications is dealing with protein flexibility. When the protein structures differ by discrete, localized changes (protonation state modifications, alterations of side chain rotamer states), it should be possible to examine the variations in protein structure directly in a single docking run with suitable algorithms, thus saving considerable computational effort. When there is a larger perturbation of protein structure, however, as in cases such as 1kv2 in which there is a significant induced-fit component of ligand binding that results in substantial backbone or loop movement, other approaches are needed. The simplest approach is to carry out flexible docking into multiple rigid protein structures and then to combine the screening results. When knowledge of the appropriate ensemble of protein structures is available, this strategy is likely to succeed. Examples of this approach will be described in a subsequent paper in the context of Extra-Precision Glide.

Ultimately, accurate molecular mechanics modeling of the protein structure will be needed to enumerate the variations in active-site geometry that can be accessed at relatively low energies. Calculations along these lines, if successful, would obviate the need for cocrystallized examples. While modeling of this type is clearly quite difficult at present, methods using continuum solvation models such as those developed by Schrödinger in principle can address this problem effectively. If this can be accomplished, it would greatly enhance the effectiveness of any docking methodology in a wide range of practical applications. Efforts along these lines are

currently underway at Schrödinger and have yielded promising early results.

Acknowledgment. This work was supported in part by grants to R.A.F. from the NIH (Grants P41 RR06892 and GM 52018). We thank Dr. Didier Rognan for providing electronic copies of receptor, ligand, and decoy data sets for the thymidine kinase and estrogen receptors and of the rank orders found for GOLD, FlexX, and DOCK.

Supporting Information Available: Tables S1−S15, listing ranks of the active ligands, their GlideScore values, their hydrogen-bonding scores, and their Coluomb−vdW interaction energies with the protein site for the 15 database screens considered in this paper. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem*. **2004**, *47*, 1739−1749.

(2) Schrödinger, L.L.C., New York.

(3) Jones, G.; Wilett, P.; Glem, R. C.; Leach, A. R.; Taylor, R. Development and validation of a generic algorithm and an empirical binding free energy function. *J. Mol. Biol*. **1997**, *267*, 727−748.

(4) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A. A fast flexible docking method using an incremental construction algorithm. *Chem. Biol*. **1996**, *261*, 470−489.

(5) Rognan, D. Bioinformatic Group, UMR CNRS, France. Personal communication to T.A.H.

(6) Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem*. **2000**, *43*, 4759−4767.

(7) Engh, R. A.; Brandstetter, H.; Sucher, G.; Eichinger, A.; Bauman, U.; Bode, W.; Huber, R.; Poll, T.; Rudolph, R.; van der Saal, W. Enzyme flexibility, solvent, and "weak" interactions characterize thrombin−ligand interactions: implications for drug design. *Structure* **1996**, *4*, 1353−1362.

(8) von der Saal, W.; Kucznierz, R.; Leinart, H.; Engh, R. A. Derivatives of 4-amino-pyridine as selective thrombin inhibitors. *Bioorg. Med. Chem. Lett*. **1997**, *7*, 1283−1288.

(9) Pearlman, D. A.; Charifson, P. S. Improved scoring of ligand−protein interactions using OWFEG free energy grids. *J. Med. Chem*. **2001**, *44*, 502−511.

(10) The geometric mean defines the average enrichment factor as the *n*th root of the product of the *n* individual enrichment factors. This definition was further modified by replacing any enrichment factor of less than 1 by 1 and by weighting the contributions such that the composite weight is the same for each receptor type, even when two or three receptor site geometries are used.

(11) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus scoring: A method of obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem*. **1999**, *42*, 5100−5109.

(12) Jain, A. N. Surflex: fully automatic flexible molecular docking using a molecular-similarity-based search engine. *J. Med. Chem*. **2003**, *46*, 499−511.

(13) Murphy, R. B.; Friesner, R. A.; Halgren, T. A. Unpublished results.

(14) Stahl, M.; Rarey, M. Detailed analysis of scoring functions for virtual screening. *J. Med. Chem*. **2001**, *44*, 1035−1042.

(15) Pargellis, C.; Tong, L.; Churchill, L.; Cirillo, P. F.; Gilmore, T.; Graham, A. G.; Grob, P. M.; Hickey, E. R.; Moss, N.; Pav, S.; Regan, J. Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. *Nat. Struct. Biol*. **2002**, *9*, 268.

(16) Whittaker, M.; Floyd, C. D.; Brown, P.; Gearing, A. J. H. Design and therapeutic application of matrix metalloproteinase inhibitors. *Chem. Rev*. **1999**, *99*, 2735−2776.

(17) Babine, R. E.; Bender, S. L. Molecular recognition of protein−ligand complexes: Applications to drug design. *Chem. Rev*. **1997**, *97*, 1359−1472.

(18) Bell, I. M.; Gallicchio, S. N.; Abrams, M.; Beese, L. S.; Beshore, D. C.; Bhimnathwala, H.; Bogusky, M. J.; Buser, C. A.; Culberson, J. C.; Davide, J.; Ellis-Hutchings, M.; Fernandes, C.; Gibbs, J. B.; Graham, S. L.; Hamilton, K. A.; Hartman, G. D.; Heimbrook, D. C.; Homnick, C. F.; Huber, H. E.; Huff, J. R.; Kassahun, K.; et al. 3-Aminopyrrolidinone Farnesyltransferase inhibitors: design of macrocyclic compounds with improved pharmacokinetics and excellent cell potency. *J. Med. Chem*. **2002**, *45*, 2388−2409.

(19) Schulz-Gasch, T.; Stahl, M. Binding site characteristics in structure-based virtual screening. Evaluation of current docking tools. *J. Mol. Model*. **2003**, *9*, 47−57.

(20) Ewing, T. J. A.; Kuntz, I. D. Critical evaluation of search algorithms for automated molecular docking and database screening. *J. Comput. Chem*. **1997**, *18*, 1175−1189.

(21) Lorber, D. M.; Shoichet, B. K. Flexible ligand docking using conformational ensembles. *Protein Sci*. **1998**, *7*, 938−950.

(22) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions. I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des*. **1997**, *11*, 425−445.

(23) Shoichet, B. K.; Kuntz, I. D. Matching chemistry and shape in molecular docking. *Protein Eng*. **1993**, *6*, 723−732.

(24) Muegge, I.; Martin, Y. C. A. A general and fast scoring function for protein−ligand interactions: a simplified potential approach. *J. Med. Chem*. **1999**, *42*, 791−804.

(25) Rognan, D.; Laumoeller, S. L.; Holm, A.; Buus, S.; Tschinke, V. Predicting binding affinities of protein ligands from three-dimensional coordinates: Application to peptide binding to class I major histocompatibility proteins. *J. Med. Chem*. **1999**, *42*, 6450−4658.

(26) Wang, R.; Liu, L.; Tang, Y. SCORE: a new empirical method for estimating the binding affinity of a protein−ligand complex. *J. Mol. Model*. **1998**, *4*, 379−384.

JM030644S

Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein−Ligand Complexes

Richard A. Friesner,* Robert B. Murphy,† Matthew P. Repasky,† Leah L. Frye,‡ Jeremy R. Greenwood,† Thomas A. Halgren,† Paul C. Sanschagrin,† and Daniel T. Mainz†

Department of Chemistry, Columbia University, New York, New York 10027, Schrödinger, Limited Liability Company, 120 West 45th Street, New York, New York 10036, Schrödinger, Limited Liability Company, 101 SW Main Street, Portland, Oregon 97204

Received December 16, 2005

A novel scoring function to estimate protein−ligand binding affinities has been developed and implemented as the Glide 4.0 XP scoring function and docking protocol. In addition to unique water desolvation energy terms, protein−ligand structural motifs leading to enhanced binding affinity are included: (1) hydrophobic enclosure where groups of lipophilic ligand atoms are enclosed on opposite faces by lipophilic protein atoms, (2) neutral−neutral single or correlated hydrogen bonds in a hydrophobically enclosed environment, and (3) five categories of charged−charged hydrogen bonds. The XP scoring function and docking protocol have been developed to reproduce experimental binding affinities for a set of 198 complexes (RMSDs of 2.26 and 1.73 kcal/mol over all and well-docked ligands, respectively) and to yield quality enrichments for a set of fifteen screens of pharmaceutical importance. Enrichment results demonstrate the importance of the novel XP molecular recognition and water scoring in separating active and inactive ligands and avoiding false positives.

1. Introduction

In two previous papers[1,2] we have described the Glide high throughput docking program and provided performance benchmarks for docking and scoring capabilities. These results have established Glide as a competitive methodology in both areas.[2−5] However, it is clear from enrichment results (ref 2) that there remains substantial room for improvement in separating "active" from "inactive" compounds. In this paper we outline and present results obtained from significantly enhanced sampling methods and scoring functions, hereafter collectively referred to as "extra-precision" (XP) Glide. The key novel features characterizing XP Glide scoring are (1) the application of large desolvation penalties to both ligand and protein polar and charged groups in appropriate cases and (2) the identification of specific structural motifs that provide exceptionally large contributions to enhanced binding affinity. Accurate assignment of these desolvation penalties and molecular recognition motifs requires an expanded sampling methodology for optimal performance. Thus, XP Glide represents a single, coherent approach in which the sampling algorithms and the scoring function have been optimized simultaneously.

The goal of the XP Glide methodology is to semiquantitatively rank the ability of candidate ligands to bind to *a specified conformation of the protein receptor*. Because of the rigid receptor approximation utilized in Glide and other high throughput docking programs, ligands that exhibit significant steric clashes with the specified receptor conformation cannot be expected to achieve good scores, even if they in reality bind effectively to an alternative conformation of the same receptor. Such ligands may be thought of as unable to "fit" into that specified conformation of the protein. For docking protocols to function effectively within the rigid-receptor approximation, some ability to deviate from the restrictions of the hard wall van der Waals potential of the receptor conformation used in docking must be built into the potential energy function employed to predict the ligand binding mode. In XP and SP Glide, this is accomplished by scaling the van der Waals radii of nonpolar protein and/or ligand atoms; scaling the vdW radii effectively introduces a modest "induced fit" effect. However, it is clear that there are many cases in which a reasonable degree of scaling will not enable the ligand to be docked correctly. For example, a side chain in a rotamer state that is very different from that of the native protein−ligand complex may block the ligand atoms from occupying their preferred location in the binding pocket. There will always be borderline situations, but in practice we have found it possible to classify the great majority of cases in cross-docking experiments as either "fitting" or "not fitting". The former are expected to be properly ranked by XP Glide (within the limitation of noise in the scoring function), while the latter require an induced-fit protocol[6,7] to correctly assess their binding affinity.[4] In the present paper, we focus on complexes where the ligand fits appropriately into the receptor, as judged by two factors: (1) the ability to make key hydrogen bonding and hydrophobic contacts and (2) the ability to achieve a reasonable root-mean-square deviation (RMSD), as compared to the native complex or as obtained by analogy with the native complex of a related ligand. Comparison by analogy is often necessary when dealing with a large dataset of active ligands, only a few of which may have available crystal structures.

Our discussion of XP Glide is divided into four different sections. First, in section 2, we describe the novel terms leading to enhanced binding affinity that have been introduced to account for our observations with regard to protein−ligand binding in a wide range of systems. The origin of these terms lies in the theoretical physical chemistry of protein−ligand interactions; however, developing heuristic mathematical representations that can be used effectively in an empirical scoring function, taking into account imperfections in structures due to the rigid receptor approximation and/or limitations of the docking algorithm, requires extensive analysis of, and fitting

* To whom correspondence should be addressed. Phone: 212-854-7606. Fax: 212-854-7454. E-mail: rich@chem.columbia.edu.
† Schrödinger, L.L.C., NY.
‡ Schrödinger, L.L.C., OR.

10.1021/jm051256o CCC: $33.50
Published on Web 09/23/2006

to, experimental data. Key aspects of this analysis, along with illustrative examples, are provided in section 2 in an effort to provide physical insight as well as formal justification for the model. In developing XP Glide, we have attempted to identify the principal driving forces and structural motifs for achieving significant binding affinity contributions with specific protein−ligand interactions, above and beyond the generic terms that have appeared repeatedly in prior scoring functions. We have found that a relatively small number of such motifs are dominant over a wide range of test cases; the ability to automatically recognize these motifs, and assign binding affinity contributions, potentially represents an advance in the modeling of protein−ligand interactions based on an empirical scheme.

In section 3, we evaluate the performance of our methodology in self-docking, with regard to both the ability to generate the correct binding mode of the complex and the prediction of binding affinity, using docked XP structures for the complexes. In section 4, the performance of the scoring function in enrichment studies (ability to rank known active compounds ahead of random database ligands) for a substantial number of targets, containing qualitatively different types of active sites, is investigated. Our treatment of the data differs significantly from what has generally prevailed in previous papers in the literature; in evaluating scoring accuracy, we distinguish cases where there are significant errors in structural prediction, as opposed to systems where the structural prediction is reasonably good, but the scoring function fails to assign the appropriate binding affinity. By using only well-docked structures to parametrize and assess scoring functions, a way forward toward a globally accurate method, in which multiple structures are employed in docking and/or induced fit methods are utilized to directly incorporate protein flexibility, is facilitated.

The parameterization of XP Glide is carried out using a large and diverse training set comprising 15 different receptors and between 4 and 106 well-docked ligands per receptor. A separately developed test set incorporating four new receptors, and additional ligands for two receptors already in the training set, is also defined. All of the receptor and ligand data is publicly available (as is our decoy set, which has been posted on the Schrodinger Web site and is freely available for downloading) and we provide extensive references documenting the origin of each ligand. The results reported below have been obtained with the Glide 4.0 release.

The development of data sets suitable for the analysis described above is highly labor intensive; consequently, our current test set is too small to draw robust conclusions, and the results reported herein must be regarded as preliminary. While the test set results are encouraging with regard to demonstration of a respectable degree of transferability, a rigorous assessment of the performance to be expected on a novel receptor will have to be performed in future publications. Nevertheless, qualitative and consistent improvement in the results for both training and test set, at least as compared to the alternative scoring functions available in Glide, is demonstrated. Finally, in the conclusion, we summarize our results and discuss future directions.

2. Glide XP Scoring Function

The major potential contributors to protein−ligand binding affinity can readily be enumerated as follows:

(1) Displacement of Waters by the Ligand from "Hydrophobic Regions" of the Protein Active Site. Displacement of these waters into the bulk by a suitably designed ligand group will lower the overall free energy of the system. Waters in such regions may not be able to make the full complement of hydrogen bonds that would be available in solution. There are also entropic considerations; if a water molecule is restricted in mobility in the protein cavity, release into solvent via ligand-induced displacement will result in an entropy gain. As one ligand releases many water molecules, this term will contribute favorably to the free energy. Replacement of a water molecule by a hydrophobic group of the ligand retains favorable van der Waals interactions, while eliminating issues concerning the availability of hydrogen bonds. Transfer of a hydrophobic moiety on the ligand from solvent exposure to a hydrophobic pocket can also contribute favorably to binding by withdrawing said hydrophobic group from the bulk solution.

(2) Protein−Ligand Hydrogen-Bonding Interactions, as well as Other Strong Electrostatic Interactions Such as Salt Bridges. In making these interactions, the ligand displaces waters in the protein cavity, which can lead to favorable entropic terms of the type discussed above in (1). Contributions to binding affinity (favorable or unfavorable) will also depend on the quality and type of hydrogen bonds formed, net electrostatic interaction energies (possibly including long range effects, although these generally are considered small and typically are neglected in empirical scoring functions), and specialized features of the hydrogen-bonding geometry, such as bidendate salt bridge formation by groups such as carboxylates or guanidium ions. Finally, differences in the interactions of the displaced waters, as compared to the ligand groups replacing them, with the protein environment proximate to the hydrogen bond, can have a major effect on binding affinity, as is discussed in greater detail below.

(3) Desolvation Effects. Polar or charged groups of either the ligand or protein that formerly were exposed to solvent may become desolvated by being placed in contact with groups to which they cannot hydrogen bond effectively. In contrast to the two terms described above, such effects can only reduce binding affinity.

(4) Entropic Effects Due to the Restriction on Binding of the Motion of Flexible Protein or Ligand Groups. The largest contributions are due to restriction of ligand translational/orientational motion and protein and ligand torsions, but modification of vibrational entropies can also contribute. As in the case of desolvation terms, such effects will serve exclusively to reduce binding affinity.

(5) Metal−Ligand Interactions. Specialized terms are needed to describe the interaction of the ligand with metal ions. We shall defer the discussion of metal-specific parameterization to another publication, as this is a complex subject in its own right, requiring considerable effort to treat in a robust fashion.

A large number of empirical scoring functions for predicting protein−ligand binding affinities have been developed.[8−19] While differing somewhat in detail, these scoring functions are broadly similar. A representative example, the ChemScore[8] scoring function, is discussed in our comments below, though similar comments would apply to many of the other scoring functions cited in refs 8−19. We briefly summarize how ChemScore treats the first four potential contributors to the binding affinity presented above:

(1) ChemScore[8] contains a hydrophobic atom−atom pair energy term of the form

$$E_{\text{phobic\_pair}} = \sum_{ij} f(r_{ij}) \tag{1}$$

Here, i and j refer to lipophilic atoms, generally carbon, and $f(r_{ij})$ is a linear function of the interatomic distance, r_{ij}. For r_{ij} less than the sum of the atomic vdW radii plus 0.5 Å, f is 1.0.

Between this value and the sum of atomic vdW radii plus 3.0 Å, *f* ramps linearly from 1.0 to zero. Beyond the sum of atomic vdW radii plus 3.0 Å, *f* is assigned a value of zero.

This term heuristically represents the displacement of waters from hydrophobic regions by lipophilic ligand atoms. Numerous close contacts between the lipophilic ligand and protein atoms indicate that poorly solvated waters have been displaced by lipophilic atoms of the ligand that themselves were previously exposed to water. The resulting segregation of lipophilic atoms, and concomitant release of waters from the active site, lowers the free energy via the hydrophobic effect, which is approximately captured by the pair scoring function above. Terms based on contact of the hydrophobic surface area of the protein and ligand, while differing in details, essentially measure the same free energy change and have a similar physical and mathematical basis.

Various parameterizations of the atom−atom pair term have been attempted, including efforts such as PLP,[9] in which every pair of atom types is assigned a different empirical pair potential. However, it is unclear whether this more detailed parameterization yields increased accuracy in predicting binding affinities. A key issue is whether a correct description of the hydrophobic effect can be achieved in all cases by using a linearly additive, pairwise decomposable functional form.

(2) ChemScore evaluates protein−ligand hydrogen-bond quality based on geometric criteria, but otherwise does not distinguish between different types of hydrogen bonds or among the differing protein environments in which those hydrogen bonds are embedded.

(3) ChemScore does not treat desolvation effects.

(4) ChemScore uses a simple rotatable-bond term to treat conformation entropy effects arising from restricted motion of the ligand.

The new XP Glide scoring function starts from the "standard" terms discussed above, though the functional form of the first three terms have been significantly revised and the parameterization of all terms is specific to our scoring function. In the remainder of this section, the functional form and physical rationale for the novel scoring terms we have developed are described with examples from pharmaceutically relevant test cases provided to illustrate how the various terms arise from consideration of the underlying physical theory and experimental data.

Form of the XP Glide Scoring Function. The XP Glide scoring function is presented in eq 2. The principal terms that favor binding are presented in eq 3, while those that hinder binding are presented in eq 4. A description of each of the following terms besides $E_{\mathrm{hb\_pair}}$ and $E_{\mathrm{phobic\_pair}}$, which are standard ChemScore-like hydrogen bond and lipophilic pair terms, respectively, follows.

$$\text{XP GlideScore} = E_{\mathrm{coul}} + E_{\mathrm{vdW}} + E_{\mathrm{bind}} + E_{\mathrm{penalty}} \quad (2)$$

$$E_{\mathrm{bind}} = E_{\mathrm{hyd\_enclosure}} + E_{\mathrm{hb\_nn\_motif}} + E_{\mathrm{hb\_cc\_motif}} + E_{\mathrm{PI}} + E_{\mathrm{hb\_pair}} + E_{\mathrm{phobic\_pair}} \quad (3)$$

$$E_{\mathrm{penalty}} = E_{\mathrm{desolv}} + E_{\mathrm{ligand\_strain}} \quad (4)$$

Improved Model of Hydrophobic Interactions: Hydrophobic Enclosure ($E_{\mathrm{hyd\_enclosure}}$). The ChemScore atom−atom pair function, $E_{\mathrm{phobic\_pair}}$ described above, assigns scores to lipophilic ligand atoms based on summation over a pair function, each term of which depends on the interatomic distance between a ligand atom and a neighboring lipophilic protein atom. This clearly captures a significant component of the physics of the hydrophobic component of ligand binding. It is assumed that the displacement of water molecules from areas with many proximal lipophilic protein atoms will result in lower free energy than displacement from areas with fewer such atoms. As a crude example, it is clear that if the ligand is placed in an active-site cavity, as opposed to on the surface of the protein, the lipophilic atoms of the ligand are likely to receive better scores. If they are located in a "hydrophobic pocket" of the protein, scores should be better than in a location surrounded primarily by polar or charged groups. Furthermore, these improved scores are likely to be correlated with improvements in ligand binding affinity.

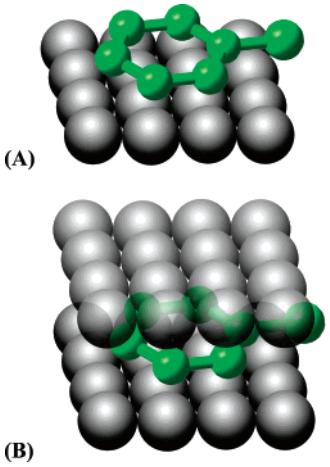

Figure 1. Schematic of a ligand group interacting with two distinct hydrophobic environments: above a hydrophobic "plane" (A) and enclosed in a hydrophobic cavity (B).

However, a function dependent only on the sum of interatomic pair functions is potentially inadequately sensitive to details of the local geometry of the lipophilic protein atoms relative to the ligand lipophilic atom in question. As an example, consider the two model distributions shown in Figure 1. In one case (A), a lipophilic ligand group is placed at a hydrophobic "wall" with lipophilic protein atoms on only a single face of the hydrophobic group. In the second case (B), the lipophilic ligand group is placed into a tight pocket, with lipophilic protein atoms contacting the two faces of the ligand group. As suggested above, one would normally expect a larger contribution to binding in the second case than in the first. However, this does not fully settle the question, which at root is whether the atom−atom pair contribution for a given ligand-atom/protein-atom distance should be identical when the ligand atom is enclosed by protein hydrophobic atoms, as opposed to when it is not, or whether there can be expected to be nonadditive effects.

From a rigorous point of view, the answer depends principally upon the free energy to be gained by displacing a water molecule at a given location. This in turn depends on how successfully that water molecule is able to satisfy its hydrogen-bonding requirements at that location, while retaining orientational flexibility. In the extreme case in which a single water molecule is placed in a protein cavity that can accommodate only one water molecule and is surrounded on all sides by lipophilic atoms that cannot make hydrogen bonds, the enthalpy gain of transferring the water to bulk solution is enormously favorable. In such a case it is not clear that a water molecule would occupy such a cavity in preference to leaving a vacuum, despite the statistical terms favoring occupancy. However, this is a rare situation not particularly relevant to the binding of a large ligand, whereas structural motifs similar to the examples in Figure 1 are quite common.

There have been a large number of papers in the literature studying, via molecular dynamics simulations, the behavior of

water in contact with various types of hydrophobic structures, including flat and curved surfaces,[20−22] parallel plates,[23,24] nanotubes,[25] and recently more realistic systems such as the hydrophobic surfaces of a protein or the interface between two protein domains.[26−28] There have also been attempts to develop general theories as to how the hydrophobic effect depends on the size and shape of the hydrophobic structure presented to the water molecules.[25,29]

A number of concepts that are clearly related to the proposals in the present paper have emerged from this work: evacuation of water (dewetting), under the appropriate conditions, from regions between two predominantly hydrophobic surfaces[1,2,9] and a model for the curvature dependence of the hydrophobic energy in which concave regions are argued to have greater hydrophobicity than convex ones.[29] However, while this work provides useful ideas and general background, development of a scoring function that can be used to quantitatively predict protein−ligand binding in the highly heterogeneous and complex environment of a protein active site requires direct engagement with a critical mass of experimental data as well as extensive parameterization and investigation of a variety of specific functional forms. In what follows, we describe the results of our investigations along these lines.

A large number of computational experiments involving modifications of the hydrophobic scoring term designed to discriminate between different geometrical protein environments have been performed. The criterion for success in these experiments is the ability of any proposed new term to fit a wide range of experimental binding free energy data and yield good predictions in enrichment studies. Key findings are summarized as follows:

(1) Ligand hydrophobic atoms must be considered in groups, as opposed to individually. The free energy of water molecules in the protein cavity is adversely affected beyond the norm primarily when placed in an enclosed hydrophobic microenvironment that extends over the dimension of several atoms. If there are individual isolated hydrophobic contacts, the water will typically be able to make its complement of hydrogen bonds anyway by partnering with neighboring waters as in clathrate structures surrounding small hydrocarbons in water.[28] After empirical experimentation, the minimum group size of connected ligand lipophilic atoms has been set at three.

(2) When a group of lipophilic ligand atoms is enclosed on two sides (at a 180 degree angle) by lipophilic protein atoms, this type of structure contributes to the binding free energy beyond what is encoded in the atom−atom pair term. We refer to this situation as *hydrophobic enclosure* of the ligand. There is some analogy here to the parallel plate, nanotube (with some sets of parameters), and protein systems in which dewetting has been observed, although the length scale of the region under consideration is smaller and (likely) more heterogeneous. The pair hydrophobic term in eq 1 is generally fit to data from a wide range of experimental protein−ligand complexes. As such, it represents the behavior of individual lipophilic ligand atoms in an "average" environment. Our new terms utilize specific molecular recognition motifs and are designed to capture deviations from this average that lead to substantial increases in potency for lipophilic ligand groups of types that are typically targeted in medicinal chemistry optimization programs. That is, placing an appropriate hydrophobic ligand group within the specified protein region leads to substantial increases in potency. Indeed, the data enabling development of this term was primarily obtained from a wide range of published medicinal chemistry efforts that provided examples of lipophilic groups that yielded exceptional increases in potency, as well as those yielding minimal increases. Our objective has been to explain these results on the basis of physical chemical principles and to develop empirical scoring terms that captured the essential physics while rejecting false positives, even with imperfect docking and the neglect of induced fit effects.

Calculation of the hydrophobic enclosure score, $E_{hyd\_enclosure}$, is summarized below with a more detailed description of the algorithm provided in Supporting Information:

(1) Lipophilic protein atoms near the surface of the active site and lipophilic ligand atoms are divided into connected groups. There are a set of rules specifying which atoms count as lipophilic and what delimits a group.

(2) For each atom in a group on the ligand, lipophilic protein atoms are enumerated at various distances.

(3) For each lipophilic ligand atom, the closest lipophilic protein atom is selected and a vector is drawn between it and the ligand atom. This is the protein "anchor" atom for that ligand atom. Vectors for all other suitably close lipophilic protein atoms are drawn to the ligand atom and their angles with the anchor-atom vector are determined. To be considered on the "opposite side" of the anchor atom, the angle between vectors must exceed a cutoff value that depends on the pair distance, with shorter distances requiring that the angle be closer to 180°. If the angle is close to zero degrees, the atom is on the "same side", and is at right angles to the anchor if the angle is close to 90°. When the angle between lipophilic protein atoms is close to 180°, we have argued this leads to an especially poor environment for waters.

(4) Each lipophilic ligand atom is assigned a score based on the number of total lipophilic contacts with protein atoms, weighted by the angle term. If no protein atom is greater than 90 degrees from the anchor atom, the angle term is zero and the atom contributes zero to the group's $E_{hyd\_enclosure}$ term. The overall score for a group is the sum over all atoms in that group of the product of the angular factor and a distance dependent factor.

(5) If the score for any ligand group is greater than 4.5 kcal/mol, the penalty is capped at 4.5 kcal/mol. This was an empirical determination based on investigating many test cases and comparing the results with experimental data. The capping is rationalized by arguing that if a very large region of this type leads to a score greater than 4.5 kcal/mol, there is probably some ability of the water molecules to compensate by interacting with each other.

An experimentally validated example of the gain in binding affinity from placing a large hydrophobic group in a pocket in which lipophilic protein atoms are present on both sides of the pocket (rings in both cases) is shown in Figure 2. Here, replacing a phenyl substituent with a naphthyl group was shown[30] to result in a 21-fold improvement in experimentally measured affinity (K_d). The naphthyl is required to fully occupy the hydrophobic pocket depicted in Figure 2.

As indicated above, the surrounding of ligand lipophilic atoms or groups by lipophilic protein atoms is referred to as hydrophobic enclosure. Our contention, here and in much of the following discussion of hydrogen bonding, is that proper treatment of hydrophobic enclosure is the key to discrimination of highly and weakly potent binding motifs and compounds. The underlying mathematical framework for describing enclosure, discussed above, could be cast in other forms, but the essential idea would remain unchanged. Detailed optimization of the numerical criteria for recognizing enclosure, and assigning

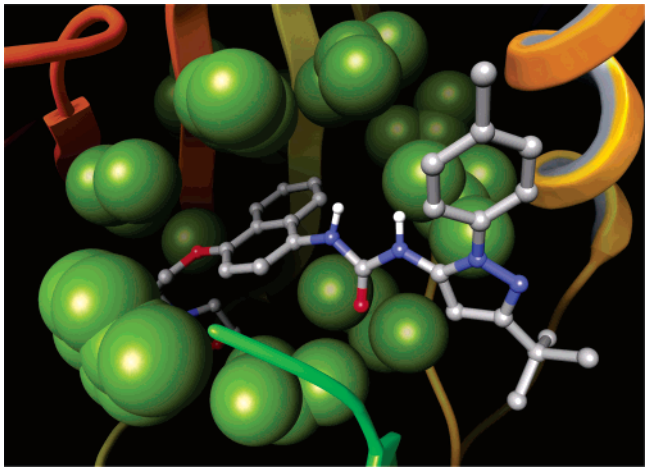

Figure 2. Boehringer active for 1kv2 bound to human p38 map kinase. The naphthyl group receives a −4.5 kcal/mol hydrophobic enclosure packing reward.

a specific contribution to the binding affinity for each motif is vital to developing methods with predictive capability.

Improved Model of Protein-Ligand Hydrogen Bonding. In developing a refined model of hydrogen bonding, we divide hydrogen bonds into three types, neutral−neutral, neutral−charged, and charged−charged. The analysis of each type of hydrogen bonding is different due to issues associated with the long-range solvation energy (Born energy) of charged groups. An initial step is to assign different default values (assuming optimal geometric features) to each of the three types of hydrogen bonds. The default values assigned are neutral−neutral, 1.0 kcal/mol, neutral−charged, 0.5 kcal/mol, and charged−charged, 0.0 kcal/mol. These assignments are based on a combination of physical reasoning and empirical observation from fitting to reported binding affinities of a wide range of PDB complexes.

The rationale for rewarding protein−ligand hydrogen bonds at all is subtle, because any such hydrogen bonds are replacing hydrogen bonds that the protein and ligand make with water. At best the net number of total hydrogen bonds on average will remain the same in the bound complex as compared to solution. However, the liberation of waters to the bulk can be argued to result in an increase in entropy, and liberation of waters around a polar protein group requires that a protein−ligand hydrogen bond with similar strength be made for a desolvation penalty to be avoided. This analysis is most plausible when both groups are neutral. The formation of a salt bridge between protein and ligand involves very different types of hydrogen bonding from what is found in solution. The thermodynamics of salt bridge formation in proteins has been studied extensively, both theoretically and experimentally,[31−34] and depends on many factors such as the degree of solvent exposure of the groups involved in the salt bridge. The default value of zero that we assign is based on the presence of many protein−ligand complexes in the PDB with very low binding affinities in which solvent-exposed protein−ligand salt bridges are formed. Assigning the contributions of these salt bridges to the binding affinity would lead to systematically worse agreement with experimental enrichment data. In XP scoring, certain features of a salt bridge are required for this type of structure to contribute to binding affinity in XP scoring. Finally, the charged−neutral default value represents an interpolation between the neutral−neutral and charged−charged value that appears to be consistent with the empirical data.

Hydrogen-bond scores are diminished from their default values as the geometry deviates from an ideal hydrogen-bonding geometry, based on both the angles between the donor and acceptor atoms and the distance. The function that we use to evaluate quality is similar to that used in ChemScore.

In what follows, specialized hydrogen-bonding motifs are described in which additional increments of binding affinity are assigned in addition to those from the ChemScore-like pairwise hydrogen-bond term. Our investigations indicate that these situations can arise for neutral−neutral or charged−charged hydrogen bonds, but not for charged−neutral hydrogen bonds. The exclusion of charged−neutral hydrogen-bond special rewards has principally been driven by our failure to date to identify motifs of this type that help to improve the agreement with experimental data. One can speculate that the lack of charge complementarity in charged−neutral hydrogen bonding precludes such structures from being major molecular recognition motifs, though further investigations with larger data sets will be needed to resolve this issue.

Special Neutral−Neutral Hydrogen-Bond Motifs ($E_{hb\_nn\_motif}$). In this section, neutral−neutral hydrogen-bonding motifs are described that were identified, based on both theoretical and empirical considerations, as making exceptional contributions to binding affinity. Such "special" hydrogen bonds represent key molecular recognition motifs that are found in many if not most pharmaceutical targets. Targeting such motifs is a central strategy in increasing the potency and specificity of medicinal compounds. Identifying such motifs through their incorporation in the scoring function should enable a dramatic improvement in both qualitative and quantitative predictions.

The critical idea in our recognition of special hydrogen bonds is to locate positions in the active-site cavity at which a water molecule forming a hydrogen bond to the protein would have particular difficulty in making its complement of *additional* hydrogen bonds. Forming such a hydrogen bond imposes nontrivial geometrical constraints on the water molecule. This is the basis for the default hydrogen-bond score, but such constraints become more problematic when the environment of the water molecule is challenging with respect to making additional hydrogen bonds such as those found in the bulk environment.

Our previous analysis of hydrophobic interactions suggests that the environment will be significantly more challenging if the water molecule has hydrophobic protein atoms on two faces, as opposed to a single face, and if few neighboring waters are available to readjust themselves to the constrained geometry of the protein−water hydrogen bond. Geometries of this type are identified using a modified version of the hydrophobic enclosure detection algorithm described previously. Replacement of such water molecules by the ligand will be particularly favorable if the donor or acceptor atom of the ligand achieves its full complement of hydrogen bonds by making the single targeted hydrogen bond with the protein group in question so that satisfaction of additional hydrogen bonds is not an issue. An example of a suitable group would be a planar nitrogen in an aromatic ring binding for example to a protein N−H backbone group. This has been observed to be essential to achieving high potency experimentally in the 1bl7 ligand binding to p38 MAP kinase, as shown in Figure 3. Here, the Met 109 hydrogen bond is known to be important for potency. Analogous hydrogen bonds have been found to be important in other kinases. In the absence of rigorous physical chemical simulations, we have used the experimental data from a significant number of diverse protein−ligand complexes to guide the development of a set of empirical rules, outlined below, for the types of ligand and

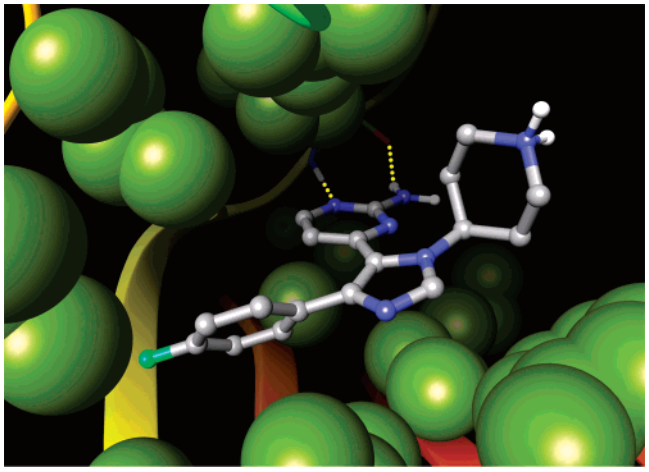

Figure 3. The 1bl7 ligand interacts with p38 MAP kinase through a neutral−neutral hydrogen bond between the ligand's aromatic nitrogen and the Met 109 N−H group.

receptor chemistries that receive this type of reward. These rules will likely evolve as more data is considered, and further simulations are undertaken.

The scoring-function term outlined above enables such hydrogen bonds to be detected automatically in advance of experimental measurement. Of equal importance, false positives, which superficially share some characteristics of the required structural motif but lack a key component, can be rejected automatically as well. Rejection of false positives has been optimized by running a given variant of the scoring function, identifying high scoring database ligands with special hydrogen bonds in locations not seen in known actives, examining the resulting structure, and altering the recognition function to eliminate the reward for such test cases.

In designing a detailed set of rules to implement the ideas outlined above, we have attempted to generalize results obtained from a wide variety of ligand−receptor systems, while at the same time avoiding false positives and respecting the basic physical chemistry principles that form the basis of the model. The detailed algorithm for detecting a single hydrogen bond in a hydrophobic environment is outlined as follows. In our implementation, the donor or acceptor atom must be in a ring with the exception of nitrogen, which is allowed to be a nonring atom. If the ligand atom is in a donor group, then all other donors of the group (e.g., the two hydrogen atoms of NH_2) must be hydrogen bonded to the protein. Only backbone protein atoms can participate in this type of special hydrogen bond. The unligated protein donor or acceptor must have fewer than three first-shell solvating waters where the waters are placed as outlined in section 3. If these criteria are met, the sum of the angular factor of the hydrophobic enclosure packing score described in this section is made over the hydrogen-bonded ligand heavy atoms and the carbon atoms attached to this ligand atom. The hydrophobic enclosure packing score used in this sum contains only the angular weight of the score described earlier and not the distance-based weight. If the absolute hydrophobic enclosure packing sum is above a defined cutoff, the hydrogen bond is considered to be in a hydrophobically constrained environment, and a special hydrogen-bond reward of 1.5 kcal/mol is applied to this hydrogen bond. In some instances, it is found that a small perturbation of the ligand can move the hydrophobic sum above the cutoff. Therefore, if the hydrophobic sum is below the cutoff, small rigid body perturbations of the ligand are made of 0.3 Å in magnitude. At each perturbed geometry, the sum is recalculated and the reward is applied if at some geometry the sum exceeds the cutoff. This procedure helps to avoid discontinuities inherent in the use of a cutoff.

Table 1.

conditions for not applying the special pair hydrogen-bond scores
ignore poor quality hydrogen bonds ($<0.05\ E_{hb\_pair}$)
ignore pairs involving the same neutral protein atom
ignore pairs involved in a salt bridge if the electrostic potential at either ligand atom is above the cutoff
ignore salt bridge pairs if either protein atom is involved in a protein−protein salt bridge
ignore ligand donor/donor pairs that come from NH_x, where $x \geq 2$ groups, the nitrogen atom is not in a ring and has no formal charge
ignore formally charged protein atoms with more than eight second-shell waters in the unligated state
ignore charged−neutral hydrogen bonds unless the protein atom is in a salt bridge
ignore pairs of different neutral acceptor atoms on the ligand
neutral hydrogen-bond pairs must satisfy ring atom and hydrophobicity environment criteria as outlined in section 2
ignore ligand hydroxyl to protein hydrogen bonds if the protein atom has zero formal charge

The situation described above identifies a structure in which a single hydrogen bond should be assigned a "special" reward. A second situation occurs when there are multiple correlated hydrogen bonds between the protein and the ligand. The physical argument is that the organization of water molecules to effectively solvate a structure of this type in the confined geometry of the active site can be even more problematic than that for the single hydrogen-bond situation described above. However, this will occur only if the waters involved in such solvation are in a challenging hydrophobic environment, with hydrophobic groups on two sides. The coupling of the special hydrogen-bond identification with the hydrophobic enclosure motif is critical if false positives are to be rejected. Correlated hydrogen bonds are routinely formed in docking with highly solvent-exposed backbone pairs, but there is no evidence from the experimental data we have examined that such structures contribute to enhanced potency.

Pair-correlated hydrogen bonds are defined as a donor/acceptor, donor/donor, or acceptor/acceptor pair of ligand atoms (referred to as "ligand atom pair") that are separated by no more than one rotatable bond (hydroxyl groups count as nonrotatable in this calculation). Several restrictions on the types of pairs that can be considered are made as detailed in Table 1. If the ligand atoms of the pair individually have zero net formal charge, they must satisfy the following hydrophobicity criterion to achieve a special hydrogen-bond reward. First, the ligand atoms must be part of the same ring or be directly connected to the same ring. Assuming the pair satisfies these restrictions, the hydrophobicity of the hydrogen-bond region is detected in a manner similar though not identical to that for a single special hydrogen bond. A sum of the hydrophobic enclosure packing score described previously is made for the pair of hydrogen-bonding ligand heavy atoms and the ring atoms directly connected to the ligand pair atoms. If a ligand atom of the pair is not a ring atom but is connected to a ring, the sum includes atoms of the ring that are nearest neighbors to the nonring ligand atom. If the absolute hydrophobic sum is above a cutoff, the hydrogen-bonded pair is given a special reward of 3 kcal/mol. Note that double counting of pair and single special hydrogen bonds is avoided by checking pairs first and excluding any rewarded hydrogen bonds found from single hydrogen-bond consideration.

Finally, the special hydrogen-bond rewards are linearly reduced with the quality of the hydrogen bond. The hydrogen-

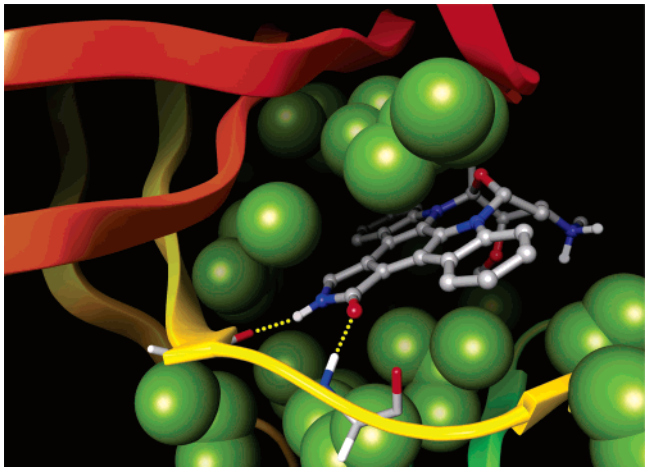

Figure 4. Staurosporine bound to human cyclin dependent kinase (CDK2). The pair of correlated hydrogen bonds receives a −3 kcal/mol reward, while the central component of the ring is given a −3 kcal/mol hydrophobic enclosure packing reward.

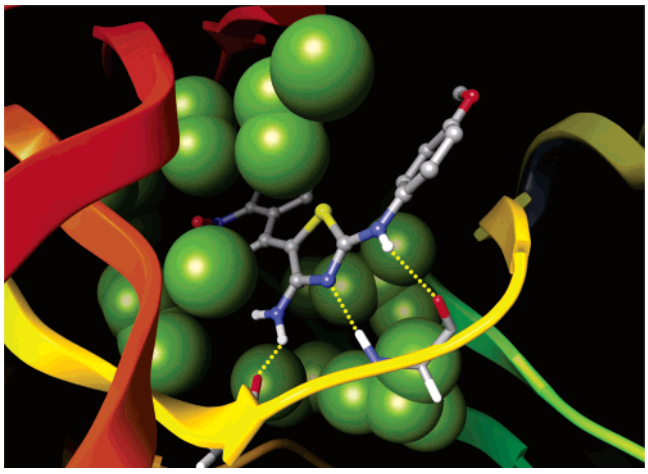

Figure 5. AG12073 bound to human cyclin dependent kinase (CDK2). The three correlated hydrogen bonds receive a −4.2 kcal/mol reward.

bond quality is measured in the sense of the pair hydrogen-bond score using the donor/acceptor distance and the angle made by the donor-heavy-atom-H vector and the H-acceptor vector. For ligand acceptor atoms in rings, the extent to which the acceptor lone pair vector is aligned with the donor-heavy-atom-H vector is evaluated. A detailed description of the algorithm for scaling the special hydrogen-bond reward is given in Supporting Information.

A substantial number of protein−ligand complexes in which motifs containing correlated hydrogen bonds that satisfy the above criteria, including the requisite hydrophobic enclosure, have been identified. A number of examples are shown below. Figure 4 depicts the 1aq1 structure of staurosporine bound to human cyclin-dependent kinase (CDK2). This type of correlated pair is also found in a number of other kinases. Some of the CDK2 actives, such as AG12073, make three correlated hydrogen bonds; this structure is shown in Figure 5. A second example is streptavidin bound to the 1stp structure of biotin (Figure 6), with three correlated hydrogen bonds in a hydrophobically enclosed region. To our knowledge, no empirical scoring function has explained the exceptionally large binding affinity of streptavidin to biotin. However, once the correlated hydrophobically enclosed hydrogen-bonding motif is recognized and assigned an appropriate score (a reward of 3 kcal/mol, consistent with other examples), the deviation between calculated and experimental binding affinity, using a docked structure, is only 0.1 kcal/mol. It should be noted that the high accuracy of this prediction is fortuitous and is not intended to suggest an ability to rigorously rank order compounds. Instead, the intent is to contrast the qualitatively reasonable prediction with that of alternative scoring functions, which typically yield results for this complex in error by ∼5−10 kcal/mol, for example, as reported in ref 35. The triply correlated, enclosed hydrogen-bonding motif also explains the low nanomolar binding affinity in the binding of fidarestat to the 1ef3 structure of aldose reductase (Figure 7) relative to a large number of ligands that achieve similarly large lipophilic scores in the highly hydrophobic active site, yet have only micromolar affinity.

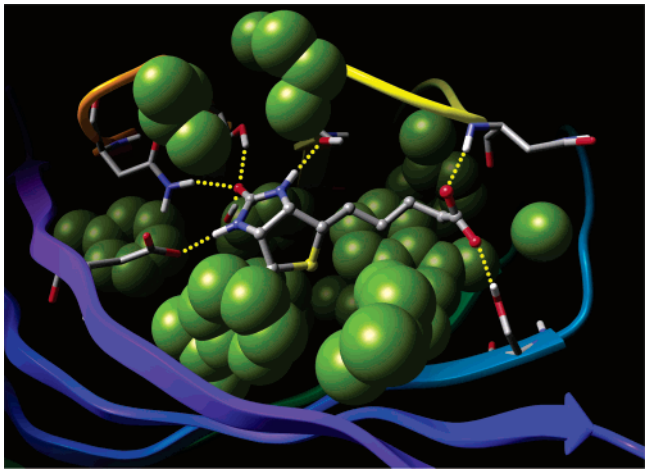

Figure 6. Biotin bound to streptavidin. The identification of a triplet of correlated hydrogen bonds in the ring in a hydrophobically enclosed region, and the three hydrogen bonds to the ligand carbonyl within that ring each contribute −3 kcal/mol rewards to this tightly bound complex (ΔG_{exp} = −18.3 kcal/mol, XP binding = −18.2 kcal/mol).

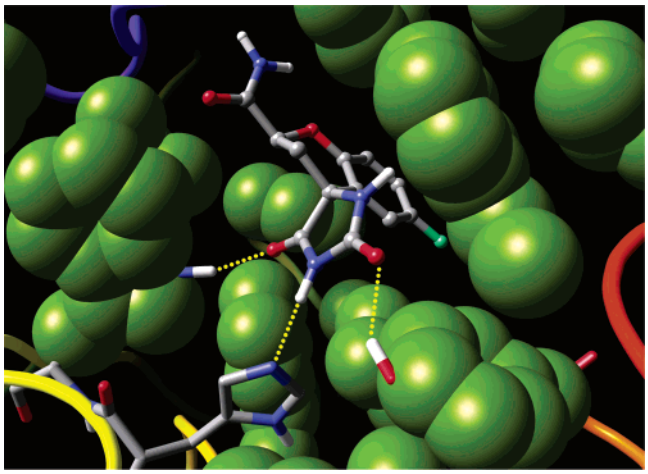

Figure 7. Fidarestat bound to aldose reductase. The triplet of special hydrogen bonds to the ring contributes −5.0 kcal/mol to the binding energy of this 9 nM inhibitor.

In our studies of various pharmaceutically relevant targets, the combination of hydrophobic enclosure with one to three correctly positioned hydrogen bonds is characteristic of every "special" neutral−neutral hydrogen-bond motif that leads to an exceptional increase in experimentally measured potency. However, additional characteristics are required to eliminate false positives. In particular, if the hydrogen-bond partner in the protein is highly solvent exposed, formation of a structure capable of solvating the group in question, while still allowing the waters involved to form a suitable number of additional

Table 2. Special Hydrogen-Bond Reward Values

hydrogen-bond moiety	reward (kcal/mol)
single bond to ring in hydrophobic environments	1.5
neutral pair in hydrophobic environments	3.0

hydrogen bonds, becomes easier. Thus, we require that the protein group(s) involved in the special hydrogen bond(s) have a limited number of waters in the first or second shell. Determination of the number of surrounding waters is carried out via the water addition code described later in this section.

The magnitude of the rewards associated with the special hydrogen bonds has been determined by optimization against a large experimental database containing a significant number of examples of each type of structure. Values are given in Table 2. It is conceivable that finer discriminations, depending upon the details of the donors and acceptors, hydrophobic environment, bound waters, and so on, could be developed, along with a correspondingly more elaborate scoring scheme. However, the present relatively simple scheme appears to work remarkably well, at least at the level of discriminating active from inactive compounds (as opposed to rank ordering, which we have not yet examined in detail) in a wide variety of test cases.

Special Charged−Charged Hydrogen-Bond Motifs ($E_{hb\_cc\_motif}$). We have identified the following features that signal enhanced binding affinity from charged−charged hydrogen bonds:

(1) The number of waters surrounding the protein component of the salt bridge. Charged groups that are fully exposed to solvent are unlikely to participate in enhanced charged−charged hydrogen bonding because the cost of displacing the solvent is simply too large. Solvent exposure is calibrated using our water scoring code, described later in this section (see E_{desolv}), by examining the number of waters in the first two shells surrounding the charged protein group.

(2) The number of charged−charged hydrogen bonds made by the charged ligand group. Three different types of salt-bridge structures have been observed: (a) Monodentate (single hydrogen bond) between the ligand group and a protein group. (b) Bidentate (two hydrogen bonds) between the ligand group and a protein group. An example of a bidentate salt bridge occurs in the 1ett structure of thrombin between a positively charged amine group and a recessed Asp 189 carboxylate in the relevant specificity pocket, as displayed in Figure 8. (c) Hydrogen bonds of one ligand group to two different protein groups. This requires having two like-charged protein groups in close proximity. This structure, which presumably creates strain energy in the apo protein, occurs with a greater frequency than might be expected. Figure 9 presents an example showing ligand Gr217029 binding to the tern N9 influenza virus of the neuramidase receptor (1bji) with a distance between carboxylate oxygen atoms of only 4.5 Å.

Empirical observations, such as the unexpectedly high potency of several neuramidase ligands including Gr217029 cited above, and physical chemical reasoning in that the electric field from the two nonsalt-bridged, proximate carboxylates is highly negative and interacts more favorably with a ligand positive charge than is typical for a salt bridge suggest that (c) provides less stabilization energy than (b), which in turn provides less stabilization energy than (a). Similarly, one would expect that a bidentate structure is more favorable electrostatically than a monodentate structure. Note, however, that unless consideration (1) is properly satisfied, none of the three structures is likely to be favorable from a free energy point of view. It is the combination of restricted water access for the protein group and an exceptionally strong electrostatic interaction between the ligand and protein group that creates the molecular recognition motif.

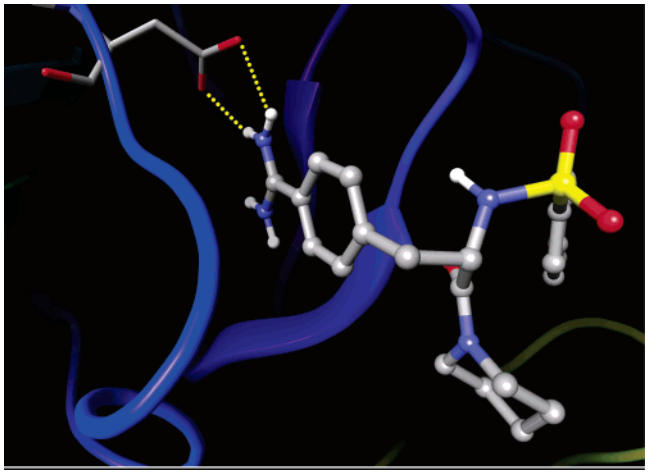

Figure 8. The bidentate hydrogen bonds in this thrombin complex bridge the ligand and Asp 189.

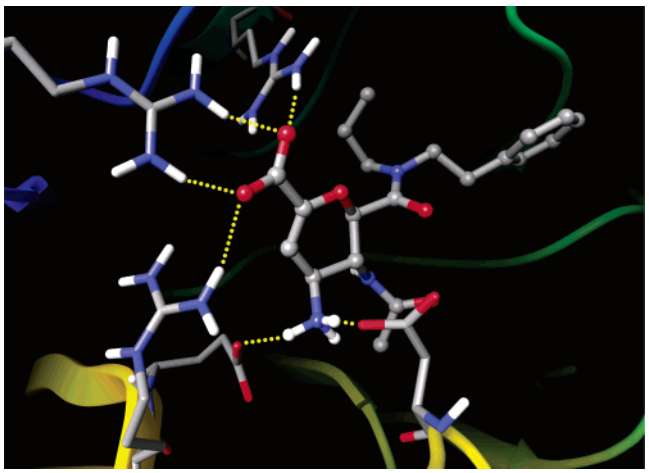

Figure 9. Gr217029 forms hydrogen bonds with two nearby Asp residues when bound to tern N9 influenza virus neuraminidase.

(3) Zwitterion ligands. A principal reason that the default value for charged−charged hydrogen bonds is set at zero is that, in forming a salt bridge, both the protein and ligand must surrender long-range contributions to the Born energy (i.e., those beyond the first shell). Satisfying the first-shell complement of hydrogen bonds is quite possible when forming a salt bridge, but the replacement of bulk water with the protein, or bound waters, clearly reduces the possible dielectric response to the ion. For a monovalent ion, the unscreened Coulomb field decreases as 1/r. Even though past the second shell dielectric screening substantially reduces further contributions, long-range effects make a nontrivial contribution to the total solvation-free energy. However, for a zwitterion, the fields from the positive and negative charges to some extent cancel at long range, yielding a dipolar field for which the second- and higher-shell contributions to the solvation free energy are significantly reduced. This cancellation depends on the separation of the two charged groups. Thus, formation of two salt bridges by the zwitterion, particularly if the two oppositely charged moieties in the ligand are spatially proximate, should be more favorable than binding a single ion. An example of a zwitterion binding in this fashion is shown in Figure 9.

Table 3. Electrostatic Rewards; Note that Double Counting Is Avoided

charge interaction	reward (kcal/mol)
charged ligand atom in low electrostatic potential environments	1.5
zwitterion configuration, range of rewards increasing with electrostatic attraction	3.0 to 4.7
positive ligand group binding to weakly solvated negative protein group	0.5
ligand CO_2- group hydrogen bound to multiple proximate positive protein residues	1.0
salt bridge pair in low solvation environment (less than nine second-shell waters about pre-ligated charge protein atom)	2.0

Table 4. 4.0 XP Binding Energies for Docking into Various GluR2 Receptors Compared to Experiment[a]

ligand	XP score	ΔG_{exp} (kcal/mol)
1ftl	−8.9	−8.3
1pwr	−12.6	−13.0
1ftj	−6.6	−8.5
1mm7	−10.7	−12.7
1mqi	−11.9	−11.9
1ftm	−11.3	−8.8
1n0t	−6.6	−5.7
1m5b	−9.3	−9.3
1m5c	−9.4	−9.3
1m5e	−6.4	−6.3

[a] These systems were used to calibrate the charged−charged hydrogen-bond recognition motif ($E_{hb\_cc\_motif}$).

(4) Cases where the ligand is positively charged and the protein is negatively charged are distinguished from those in which the charge states are reversed.

(5) Strength of the electrostatic field at the ligand. An enhanced binding affinity for a salt bridge is assigned if the site at which the ligand charge is placed is sufficiently electrostatically favorable. The electrostatic field at the ligand site is summed using constant and distance-dependent dielectric models, and cutoffs are imposed for assigning rewards based on empirical optimization over our suite of test cases. These cutoffs help reduce the number of false positives receiving special charged−charged rewards.

Table 3 enumerates the various special charged−charged rewards for motifs based on the five categories discussed above. The numerical values have been optimized based on fitting to our entire test suite.

Table 4 displays the XP active scores for a series of Glur2 receptors versus the experimental binding energies. Along with neuramididase, this is a system for which electrostatic interactions are particularly important. As such, it provides an important contribution to the training set. The good agreement displayed was achieved by using a combination of the electrostatic terms discussed above.

Other Terms. A number of other types of specialized terms have been investigated. These include terms rewarding pi stacking and pi-cation interactions (E_{PI}), rewards for halogen atoms placed in hydrophobic regions, and an empirical correction enhancing the binding affinity of smaller ligands relative to larger ones. These parameterizations were in many cases performed using limited data, and we do not view them yet as fully mature. As such, details will not be presented in the present publication. These terms are relatively small compared to the enclosure and charged−charged terms discussed above, but can have a nontrivial impact in specific cases. For example, the pi-cation term, for which a reward of 1.5 kcal/mol is assigned, is important for the acetylcholinesterase test case discussed below.

Terms that Penalize Binding in the XP Scoring Function. The most important physical effects that oppose binding are strain energy of the ligand, protein, or both, loss of entropy of ligand and protein, and desolvation of the ligand or protein. The penalty terms developed are targeted in all three areas, although terms addressing strain energy and entropic loss do not necessarily represent a significant advance as compared to previous terms described in the literature. In developing the penalty terms, some fundamental limitations arise from the rigid-receptor approximation and the use of empirical scoring functions rather than full energy expressions. One consequence is that, in our view, it is not possible to completely reject false positives with an empirical approach. However, significant improvements are possible as compared to previous efforts, as we shall demonstrate below.

Water Scoring: Rapid Docking of Explicit Waters (E_{desolv}). A number of approaches to incorporating desolvation penalties into a high throughput docking code have been presented in the literature.[35,36] However, the methods in these papers are based on continuum-solvation approaches. For computing protein−ligand binding affinities, the role of individual waters can be critical, and continuum models often provide poor results in treating bound waters in a confined cavity. Therefore, we have chosen to implement a crude explicit water model that can be rapidly evaluated yet captures the basic physics of solvation within the confines of the protein−ligand complex active-site region.

The approach employed is to use a grid-based methodology and add 2.8 Å spheres, approximating water molecules, to high-scoring docked poses emerging from the initial round of XP docking. In principle, this methodology is similar to that used in the GRID program, though, algorithmic details have been optimized to achieve speed, critical in the present application. The CPU time for water addition to a single pose is 3−8 CPU seconds (AMD Athelon MP 1800+ processor running Linux) on average depending on ligand size.

Once waters have been added, statistics are tabulated with regard to the number of waters surrounding each hydrophobic, polar, and charged group of the ligand and active site of the protein. When a polar or charged ligand or protein group is judged to be inadequately solvated, an appropriate desolvation penalty is assessed. Additionally, the environment of each active-site water is itself probed to search for cases in which waters make an unusual number of hydrophobic contacts. A penalty is assigned if the number of such contacts for an individual water molecule exceeds a given threshold. Water scoring statistics are also used to determine whether special hydrogen-bonding rewards should be assigned, as was discussed previously.

Contact Penalties (E_{lig_strain}). Penalizing strain energy in rigid-receptor docking is probably the single most difficult component of an empirical scoring function. The problem arises from the fact that, in a typical cross-docking situation, the ligand has to adjust to fit into an imperfect (from its point of view) and rigid cavity. This often requires ligands to adopt higher-energy, nonideal torsion angles. Considering the rigid-receptor approximation that is made, it is difficult to determine whether strained ligand geometries would arise if induced-fit effects were properly accounted for or whether strained ligand geometries would be a true requirement for docking to that receptor. Furthermore, even native ligand geometries have been found to exhibit high strain energies.[37] Given this limitation, the function used to penalize poses with close internal contacts is fairly lenient and only looks for severe cases of bad internal

contacts. One function simply counts the number of intramolecular heavy atom contacts below roughly 2.2 Å and rejects a pose entirely if there are more than three such contacts. A second, more sophisticated function assembles the contacts into groups and evaluates a penalty based on the size of the contacting groups, the range of contacts, and the extent to which the groups lie on the periphery of the molecule. Empirically, it has been found that peripheral groups are more difficult to penalize for intraligand contacts than are more centrally located groups.

Implementation Issues. Application of XP penalty terms, particularly those related to desolvation, imposes hurdles that make it difficult for random database ligands to achieve good scores in virtual screening. These hurdles do not exist in alternative programs that ignore desolvation effects and strain energy. On the other hand, if the terms are inaccurately defined, they will adversely affect active compounds. Furthermore, a definition that would be "accurate" for a high-resolution structure may function poorly for docked structures, particularly in the rigid-receptor framework, due to inaccuracies in sampling. This issue arises routinely in practice, such as when an active ligand could avoid a penalty by moving a few tenths of an Angstrom in some direction but is blocked by the rigid protein. Similarly, the sampling algorithm may simply fail to find the superior pose.

We have found that extensive sampling to enable ligands to avoid penalties when possible is an essential component of Glide XP scoring. If the penalties are due to limitations in the rigid-receptor approximation, the only solutions are (a) to adjust the parameters so that they allow the test case in question to escape penalization or (b) to accept that the active compound does not "fit" into the particular version of the receptor being used and to dock into multiple structures and/or employ induced-fit methodologies. For many cases, however, better sampling in the relevant phase space can locate ligand geometries that are able to avoid the penalties. The XP Glide sampling algorithm was explicitly designed with this objective in mind.

XP Glide Sampling Methodology. XP Glide sampling begins with SP Glide docking, as described in refs 1 and 2, but using a wider "docking funnel" so that a greater diversity of docked structures is obtained. For XP docking to succeed, SP docking must produce at least one structure in which a key fragment of the molecule is properly docked. This has been the case in the great majority of systems that have been investigated.

The second step in XP sampling is to assign various fragments of the molecule as "anchors" and to attempt to build a better-scoring pose for the ligand starting from each anchor. Typical anchors are rings, but can be other rigid fragments as well. Various positions of the anchors are clustered, representative members of each cluster are chosen, and the growing of side chains from relevant positions on the anchor is initiated.

The growing algorithm proceeds one side chain at a time, thereby avoiding the combinatorial explosion of total molecular conformations that occurs when all side chains are considered together. Because the anchor fragment is already positioned in the protein, most side-chain conformations can be trivially rejected based on steric clashes. The Glide "rough scoring" function is used to screen the initial side chain conformations. Most importantly, these conformations can be grown at extremely high resolution (4 degrees for each rotatable bond) because the total number of conformations considered at any one time is being constantly pruned via screening and clustering algorithms. It is this high-resolution sampling that enables difficult cross-docking cases to be effectively addressed, and that ultimately allows penalties to be avoided when possible.

After the individual side chains are grown, a set of candidate complete molecules is selected by combining high-scoring individual conformations at each position and eliminating structures with significant steric clashes between side chains. Candidate structures are minimized using the standard Glide total-energy function, which employs a distance-dependent dielectric to screen electrostatic interactions and are ranked according to the Emodel Glide pose-selection function,[1] comprising the molecular mechanics energy plus empirical scoring terms. Then, the grid-based water addition technology is applied to a subset of the top scoring structures, penalties are assessed as discussed above, and the full XP scoring function is computed.

At this stage, structures with the largest contributions from terms that promote binding may have penalties of various types. The next stage of the algorithm, critical to obtaining suitable results, is to attempt to evade penalties by regrowing specific side chains from such poses. The side chain that is the cause of the penalty can be targeted and, by focusing on this region of the ligand exclusively, a much larger number of candidate structures covering this region of phase space can be retained and minimized. The algorithm results in a significant increase in the density of poses and locates penalty-free structures when possible, despite the fact that the penalty terms are not at present encoded in the energy gradient. Finally, a single pose is selected based on a scoring function that combines weighted Coulomb/van der Waals protein−ligand interaction energies, the terms favoring binding affinity, and the various penalty terms.

XP Glide Parameterization: Philosophy and Implementation. The novel terms that we have described above have been developed via a combination of reasoning from basic physical chemistry principles and examining a large set of empirical data, as discussed further below. Because the terms are calculated via fast empirical functions (as opposed to rigorous atomistic simulations), extensive parameterization is required to obtain results in reasonable agreement with experiment. The large number of parameters employed in turn necessitates the use of a large number of examples in the training set; to avoid overfitting, the training set must be substantially larger than the number of parameters that are adjusted.

The total number of parameters in the current XP scoring function is on the order of 80; this includes parameters for desolvation penalties, hydrophobic enclosure, special neutral−neutral and charge−charge hydrogen bonds, and pi-cation and pi-stacking interactions. Parameters are required to convert various geometrical criteria into specific scores. The PDB complexes below, as well as the enrichment studies in the training set, were used to develop the parameter values. False positives, as well as known actives, were incorporated into the optimization protocol, so the data in the training set exceeds the total number of PDB complexes and known actives by a considerable margin (although a precise calculation of the total number of data points in the training set is very difficult to produce, as for example not every database ligand was competitive with known actives in ranking, and noncompetitive compounds played no role in parameter optimization.

Because the scoring function contains nonlinear functional forms, a rigorous optimization algorithm would also be nonlinear (rather than a simple least-squares fit); furthermore, constraints would be imposed based on physically reasonable values of the various parameters. The present set of parameters was in fact determined by a heuristic approach; a small number of

paradigmatic test cases were identified for each type of term, initial values were fit to yield reasonable results for these parameters, and the parameter set was then tested on the entire training set. Problem cases were then identified by these tests, and reoptimization was carried out to improve the worst outliers. A fully numerical optimization protocol would quite likely improve results for the training set, but it is unclear whether a corresponding improvement in the test set would result (note that test set results were not generated until the current parameter set was frozen in the Glide 4.0 release). As we develop a larger test set, we will investigate the use of more automated optimization algorithms, with likely quantitative improvements in the predictive capabilities of the scoring function.

3. Structural and Binding Affinity Prediction Results for PDB Complexes

A set of 268 complexes from the PDB, which we have previously used to assess the docking accuracy and scoring capabilities of Glide SP,[1] have been selected, of which 198 have reliable experimentally determined binding affinities as determined by our extensive examination of the literature for each case. This set of complexes displays a wide range of active-site cavities and protein−ligand interactions. The parameters of the scoring function were simultaneously optimized to reproduce the experimental binding affinity data and yield quality enrichment factors/binding affinities for the database screening tests that are discussed below.

An evaluation of the docking accuracy of Glide SP was presented in ref 1, and the XP results from docking MMFFs[38] optimized ligand structures shown in Table 5 are very similar. This suggests that the sources of error in docking accuracy are due to issues other than those addressed by the XP scoring-function modifications. In some cases, near symmetry in the ligand leads to docked poses that are functionally equivalent to those in the crystal structure, for example in terms of protein−ligand contacts, but that exhibit a large RMSD; such cases are identified in Table 5. A principal source of errors in pose RMSD appears to be the charge distribution of the ligand, which, in a standard force field representation, may not accurately distribute formal ionic charges and does not incorporate polarization effects. Cho and co-workers have demonstrated that low RMSD ligand poses can be reliably generated by utilizing more accurate polarized charges, where the ligand charge distribution is computed in the protein environment via QM/MM methods.[39] We may incorporate this methodology into future XP Glide releases for optional use. The quality of structural prediction shown by XP Glide is sufficient for the great majority of ligands to enable an adequate assessment of the scoring function to be made.

In our optimization protocol, we employed docked protein−ligand complexes, retaining only those with protein−ligand contacts that predominantly agree with those in crystal structures. In this fashion, we avoid corrupting the fitting process with irrelevant data such as would be provided by a grossly incorrect pose, yet include a realistic level of variation in the input structure. This is particularly critical in optimization of the penalty terms. If the sampling algorithm cannot avoid incorrect penalties in self-docking, it is unlikely to be able to do so in a much more challenging cross-docking situation. By incorporating docked poses of PDB complexes into the optimization process, the penalty function can be tuned to improve the agreement with experimental binding affinities while avoiding inappropriately penalizing active compounds, keeping in mind that there are also cases where the penalty terms are in fact appropriate.

Before discussing binding affinity predictions, a key point that has generally been neglected previously should be noted. An empirical scoring function that considers only protein−ligand interactions with no a priori information concerning the apo structure of the protein cannot, by definition, take into account the reorganization energy of the protein required to accommodate the ligand. In many cases, the protein is relatively rigid, the ligand fits without major rearrangements, and neglect of this term is acceptable. However, there are cases where it is overwhelmingly likely that the induced-fit energies are substantial. The most obvious cases are those in which an allosteric pocket is created to accommodate the ligand. This occurs, for example, in nonnucleoside reverse transcriptase inhibitors (NNRTIs) of HIV reverse transcriptase (HIV-RT) and is also manifested in the large-scale motion of the activation loop in p38 MAP kinases required to produce the DFG-out conformation to which inhibitors such as BIRB796 bind. However, there are many more subtle cases in which side chains or backbone groups alter their positions nontrivially, and this should introduce some energetic cost.

The goal of an empirical scoring function should be to predict the binding affinity of the ligand to the *structure with which it is presented.* That is, given the formal impossibility of predicting reorganization energy in any such scheme, this value should be removed from the experimental binding affinity. Otherwise, one will be fitting to an incorrect experimental number, given the goal of the exercise. A perfect scoring function of this type will correctly rank-order candidate ligands in their ability to bind to the structure at hand. If this structure is known to have a low reorganization energy, compounds with good scores when docked into that structure should yield satisfactory experimental binding affinities. For allosteric pockets and other sites with larger reorganization energies, one would expect that more favorable scores would be needed to yield the desired experimental binding affinity. The problem of comparing scores between ligands docked into different conformations of the receptor can then be treated separately. The problem is highly nontrivial, requiring either a heuristic procedure incorporating experimental information or the brute force ability to compare free energies of different protein conformations.

A qualitative observation that we have made, confirmed in a large number of examples, is that a large hydrophobic enclosure score is a signature of significant protein rearrangement and possibly creation of an allosteric pocket. Ligands binding tightly to both the HIV-RT NNRTI site and p38 DFG-out conformation, for example, generally receive maximal hydrophobic-enclosure scores. Furthermore, the total XP scores of these ligands are substantially higher in absolute terms than their experimental binding affinities would mandate. This is completely consistent with the ideas discussed above, in which the reorganization energy of the protein must be subtracted from the empirical binding affinity score to produce the correct experimental binding affinity.

In the initial Glide XP parameterization with PDB cocrystallized structures, systems with exceptionally large protein reorganization energies, as predicted by large hydrophobic enclosure contributions, have been intentionally omitted. This omission includes allosteric sites such as the HIV-RT and p38 structures mentioned above, and complexes such as staurosporine/CDK2 (PDB code 1aq1), in which the CDK2 pocket must expand substantially to accommodate the unusually large staurosporine ligand. If one believes that the scoring function is accurate, the protein reorganization energy can be inferred from the computed rigid receptor and experimental binding

Table 5. 4.0 XP and SP Binding Scores and Heavy Atom RMS (Å) Values for Docking into PDB Entries[a]

		GlideScore (kcal/mol)			RMS (Å)				GlideScore (kcal/mol)			RMS (Å)	
PDB	ΔG_{exp}	XP	XP-corr	SP	XP	SP	PDB	ΔG_{exp}	XP	XP-corr	SP	XP	SP
1aaq	−11.5	−10.6	−10.6	−11.5	2.01	1.40	1e5i		−7.4	−7.4	−11.2	0.28	0.17
1abe	−8.9	−7.8	−7.8	−8.8	0.31	0.40	1eap	−8.5	−12.6	−11.8	−9.2	0.65	2.38
1abf	−7.4	−8.0	−8.0	−9.3	0.17	0.14	1ebg	−14.8	−11.0	−11.0	−18.3	0.34	0.26
1acj	−10.0	−11.4	−9.9	−8.4	2.81	4.61	1ecv	−6.6	−7.4	−7.4	−9.4	0.24	0.18
1acm	−10.3	−12.2	−12.2	−13.1	0.40	0.32	1eed	−6.5	−12.0	−12.0	−8.3	11.29	1.58
1aco	−4.9	−2.3	−2.3	−10.1	0.34	0.29	1ejn	−7.7	−9.8	−7.8	−10.0	0.34	0.12
1add	−9.2	−10.3	−8.3	−9.9	0.83	0.70	1ela	−8.7	−8.4	−8.4	−5.1	0.39	6.01
1adf	−6.2	−7.9	−7.9	−11.8	3.03	9.92	1elb	−9.8	−6.3	−6.3	−7.0	5.42	4.29
1aha		−7.9	−7.9	−8.2	0.36	0.11	1elc	−9.4	−7.3	−7.3	−6.0	6.53	7.92
1ake		−14.0	−14.0	−3.9	15.45	14.95	1eld	−9.1	−6.5	−6.5	−4.8	0.32	3.94
1apb	−7.9	−7.7	−7.7	−9.4	0.06	0.10	1ele	−9.3	−8.0	−8.0	−6.4	0.36	0.38
1apt	−12.8	−11.9	−11.9	−11.0	1.48	1.24	1epb		−13.6	−11.1	−8.0	2.24	1.89
1apu	−10.2	−8.7	−8.7	−7.6	0.61	1.24	1eta		−3.9	−3.9	−3.5	8.69	1.85
1apv	−12.3	−11.2	−11.2	−8.5	0.63	0.48	1etr	−10.1	−11.7	−9.7	−9.4	0.71	0.68
1apw	−10.9	−11.0	−11.0	−9.0	0.97	0.32	1ets	−11.2	−13.7	−11.7	−11.7	1.32	1.44
1atl	−8.6	−10.6	−10.6	−7.8	0.87	3.47	1ett	−8.0	−12.6	−10.6	−9.5	0.62	0.58
1avd		−16.4	−16.4	−10.4	0.82	0.55	1ezq	−12.3	−11.4	−9.4	−12.6	0.75	0.21
1azm		−5.1	−5.1	−6.2	1.89	2.51	1f0r	−10.4	−13.3	−11.3	−10.1	2.11	0.59
1b6j	−10.8	−18.3	−18.3	−15.8	2.98	0.43	1f0s	−10.6	−11.3	−11.3	−9.4	2.08	0.35
1b6k	−11.9	−14.8	−12.8	−13.8	1.04	1.06	1f0t	−8.2	−7.7	−7.7	−9.8	0.38	0.24
1b6l	−11.3	−11.1	−11.1	−8.7	0.92	1.18	1f0u	−9.8	−10.8	−8.8	−10.5	1.56	1.54
1b6m	−11.5	−13.9	−11.9	−11.4	0.73	3.17	1fen		−12.8	−12.1	−8.4	1.10	0.40
1baf		−9.5	−9.5	−8.3	1.17	1.08	1fh8	−9.4	−10.7	−10.7	−10.3	0.20	0.20
1bap	−9.3	−8.2	−8.2	−9.1	0.39	0.38	1fh9	−8.8	−6.0	−6.0	−5.3	2.11	1.99
1bbp		−14.3	−12.4	−11.8	5.28	5.11	1fhd	−9.3	−8.1	−8.1	−8.7	5.37	0.46
1bkm		−11.6	−11.6	−14.6	4.77	2.36	1fjs	−13.6	−13.6	−11.6	−12.5	2.06	2.42
1bma	−6.3	−7.9	−7.9	−7.5	0.68	1.94	1fkg	−10.9	−13.1	−12.6	−8.4	1.21	1.33
1bra	−2.5	−5.5	−3.5	−7.8	2.26	0.32	1fki	−9.5	−10.2	−10.2	−7.7	1.30	1.29
1byb	−19.0	−14.2	−14.2	−11.3	0.56	0.46	1fq5	−11.5	−17.0	−17.0	−14.0	1.96	2.43
1c1b		−18.2	−15.7	−10.8	0.91	0.45	1fvt		−13.2	−13.2	−8.4	0.88	0.88
1c3i		−12.6	−12.6	−11.9	0.61	0.43	1g45	−11.8	−8.2	−8.2	−6.0	7.88	4.02
1c5p	−6.4	−6.2	−6.2	−8.6	0.27	0.25	1g46	−12.1	−8.2	−8.2	−6.3	8.06	4.50
1c83	−6.6	−7.3	−7.3	−9.9	0.17	0.14	1g48	−11.5	−7.0	−7.0	−5.9	1.88	3.77
1c84	−6.8	−7.5	−7.5	−8.8	0.26	0.32	1g4j	−11.9	−6.9	−6.9	−6.9	5.56	3.51
1c86	−6.4	−8.4	−8.4	−10.9	0.19	0.18	1g4o	−11.3	−7.6	−7.6	−6.0	3.39	4.21
1c87	−5.7	−8.1	−8.1	−10.9	0.28	0.21	1g52	−13.0	−8.2	−8.2	−5.7	8.01	4.26
1c88	−7.2	−7.5	−7.5	−11.8	0.25	0.22	1g53	−12.3	−8.4	−8.4	−6.4	7.88	4.53
1c8k		−8.1	−8.1	−7.3	3.28	5.50	1g54	−12.0	−7.2	−7.2	−5.4	8.45	5.14
1cbs	−9.8	−7.5	−7.5	−7.5	0.63	0.39	1ghb	−1.7	−3.2	−3.2	−9.7	0.45	0.30
1cbx	−8.7	−7.8	−7.8	−13.1	0.28	0.48	1glp		−4.5	−4.5	−8.1	0.75	0.32
1cde		−15.2	−15.2	−11.5	1.62	1.71	1glq		−6.2	−6.2	−9.7	0.46	0.32
1cdg	−3.3	−2.4	−2.4	−4.8	6.48	9.83	1gsp		−7.7	−7.7	−7.8	1.10	2.79
1cil	−12.9	−5.4	−5.4	−6.2	3.61	3.92	1hbv	−8.7	−12.1	−12.1	−5.5	2.19	2.07
1cnx	−10.0	−6.8	−6.8	−7.5	6.54	6.53	1hdc	−8.2	−9.8	−9.8	−8.3	0.56	0.35
1com	−5.4	−7.3	−7.3	−9.0	0.55	3.74	1hdy	−7.8	−4.8	−4.8	−4.2	1.65	1.70
1coy		−8.7	−8.7	−9.1	0.44	0.29	1hef	−12.1	−11.2	−11.2	−10.3	6.24	6.43
1ctr	−5.8	−7.5	−7.5	−5.7	2.27	2.59	1hfc	−7.5	−8.9	−8.9	−9.4	2.25	2.26
1ctt	−6.2	−6.9	−6.9	−7.0	0.60	5.04	1hgg	−3.4	−6.4	−6.4	−7.2	1.37	1.47
1d3d	−12.4	−12.7	−12.7	−10.4	1.52	2.74	1hgh	−3.9	−5.7	−5.7	−6.2	4.70	0.49
1d3p	−8.9	−9.6	−9.6	−10.3	1.72	2.05	1hgi	−3.7	−3.9	−3.9	−6.9	0.37	0.23
1d7x		−7.2	−7.2	−9.1	0.58	0.44	1hgj	−2.3	−4.8	−4.8	−4.0	0.64	0.34
1d8f		−6.8	−6.8	−6.8	4.25	4.31	1hih	−11.0	−11.8	−11.8	−11.0	1.26	1.29
1dbb	−12.3	−13.2	−13.0	−10.3	0.37	0.41	1hps	−12.6	−12.2	−12.2	−12.8	11.93*	2.09
1dbj	−10.4	−15.9	−13.4	−9.5	0.32	0.21	1hpv	−12.6	−9.3	−9.3	−10.0	1.05	0.93
1dbk	−11.0	−15.2	−12.9	−9.0	0.57	0.40	1hpx	−12.7	−11.5	−11.5	−13.1	3.34	3.31
1dbm	−12.9	−15.3	−13.1	−9.2	1.95	1.97	1hri	−5.9	−8.9	−8.9	−7.3	10.09	2.26
1dd6		−13.5	−13.5	−9.1	1.36	8.27	1hsg	−12.8	−12.5	−12.5	−13.3	0.41	0.35
1dds	−11.3	−10.8	−10.8	−7.4	1.75	2.38	1hsl	−9.8	−8.4	−6.4	−9.4	1.31	1.31
1dhf	−10.1	−8.7	−8.7	−8.5	6.31	5.62	1hte	−8.6	−9.9	−9.9	−9.6	7.32*	7.36
1did	−4.8	−3.8	−3.8	−6.6	3.27	4.15	1htf	−9.3	−10.9	−10.9	−9.4	2.19	2.74
1die	−2.9	−5.8	−5.8	−6.9	0.34	0.77	1hti	−7.0	−4.5	−4.5	−5.7	4.40	1.60
1dih	−7.8	−8.2	−8.2	−14.3	2.62	1.78	1hvr	−13.0	−13.7	−13.7	−7.8	1.60	1.75
1dm2		−13.9	−13.9	−10.3	0.66	0.69	1hyt		−8.0	−8.0	−10.7	2.65	0.43
1dog	−5.5	−6.5	−6.5	−9.5	3.77	3.77	1icn		−10.3	−7.8	−1.7	9.06	2.04
1dr1	−7.4	−7.6	−7.6	−7.0	0.37	1.46	1ida	−11.9	−13.1	−13.1	−12.2	1.95	2.12
1dwb	−4.0	−6.0	−4.0	−7.7	0.29	0.32	1igj	[b]	−9.2	−9.2	−7.0	0.72	0.46
1dwc	−10.3	−9.3	−9.3	−8.8	2.06	0.89	1imb	−5.7	−6.6	−6.6	−10.0	1.84	1.64
1dwd	−11.4	−13.2	−11.2	−11.2	0.47	1.43	1ivb		−5.2	−5.2	−7.0	3.27	0.47
1ivc		−3.4	−3.4	−5.0	2.05	1.89	1wap		−8.1	−8.1	−10.2	0.23	0.19
1ivd	−4.3	−6.3	−6.3	−6.2	0.73	0.73	1xid		−5.7	−5.7	−6.6	4.02	4.32
1ive		−3.6	−3.6	−5.6	5.11	5.17	1xie		−4.8	−4.8	−6.8	2.60	3.91
1ivf		−7.7	−7.7	−6.8	0.61	0.59	2ack		−9.3	−9.3	−7.0	1.07	0.88
1lah	−10.3	−7.3	−7.3	−9.6	0.52	0.19	2ada		−9.5	−9.5	−9.0	0.59	0.46
1lcp	−9.1	−7.7	−7.7	−8.8	1.82	1.06	2cgr	−9.9	−6.5	−6.5	−10.8	0.56	0.52
1ldm	−7.4	−6.3	−6.3	−7.3	1.34	1.35	2cht	−7.5	−7.7	−7.7	−11.4	0.48	0.51

Table 5. Continued

		GlideScore (kcal/mol)			RMS (Å)				GlideScore (kcal/mol)			RMS (Å)	
PDB	ΔG_{exp}	XP	XP-corr	SP	XP	SP	PDB	ΔG_{exp}	XP	XP-corr	SP	XP	SP
1lic		−6.1	−6.1	−5.3	3.96	4.99	2cmd	−6.2	−8.2	−8.2	−10.3	0.65	0.34
1lmo		−7.6	−7.6	−6.8	8.40	0.87	2cpp	−8.3	−8.2	−8.2	−6.8	0.15	3.04
1lna		−6.6	−6.6	−6.8	1.50	0.90	2ctc	−5.3	−7.4	−7.4	−10.1	1.43	1.58
1lst		−6.1	−4.1	−7.7	0.75	0.27	2dbl	−11.8	−12.2	−11.0	−8.8	2.40	0.81
1mbi	−2.6	−3.9	−3.9	−4.1	1.92	1.65	2gbp	−10.1	−9.0	−9.0	−12.5	0.61	0.14
1mcr	−4.3	−8.3	−8.3	−7.5	5.82	4.33	2ifb	−7.4	−8.7	−6.2	−2.4	2.27	1.77
1mdr	−5.4	−8.2	−8.2	−9.6	1.95	0.54	2lgs		−3.4	−3.4	−9.5	0.88	0.53
1mfe	−7.2	−8.0	−8.0	−7.4	6.09	1.78	2mcp	−7.1	−5.6	−5.6	−5.6	1.54	1.25
1mld		−6.6	−6.6	−10.6	0.25	0.22	2phh	−6.4	−7.9	−7.9	−8.7	0.47	0.41
1mmq	−10.3	−13.8	−13.8	−11.0	0.67	0.33	2pk4	−5.9	−8.8	−6.8	−5.8	0.65	0.85
1mnc	−12.3	−11.9	−11.9	−11.1	0.33	0.73	2plv		−14.3	−11.8	−7.3	1.78	1.90
1mrg		−7.8	−7.8	−7.8	0.15	0.12	2r04	−8.0	−10.6	−8.1	−8.9	1.44	0.75
1mrk	−6.2	−11.3	−11.3	−9.4	1.21	1.17	2r07		−10.9	−8.4	−8.7	0.92	0.67
1mup		−9.3	−7.6	−6.2	4.50	4.05	2sim	−4.7	−7.1	−7.1	−10.5	0.82	0.94
1nco	−10.6	−10.4	−10.4	−12.1	0.60	0.33	2tmn	−8.0	−10.1	−10.1	−10.4	0.65	0.50
1nis	−4.1	−4.0	−4.0	−7.5	0.26	0.45	2tpi	−5.9	−8.6	−6.6	−9.5	0.26	1.15
1nnb	−7.2	−5.5	−5.5	−8.7	1.39	0.25	2upj	−14.2	−11.1	−11.1	−10.7	3.24	2.69
1nsc	−4.1	−5.7	−5.7	−10.5	0.66	0.24	2xis	−7.9	−4.9	−4.9	−7.3	2.05	2.40
1nsd	−7.2	−6.3	−6.3	−9.0	0.74	0.22	2yhx		−4.3	−4.3	−5.6	1.91	2.19
1odw		−11.5	−11.5	−5.7	3.91	2.59	3cla	−6.7	−6.4	−6.4	−5.4	5.09	6.06
1okl	−8.2	−7.2	−7.2	−5.8	0.38	3.14	3dfr	−14.0	−15.3	−13.3	−11.3	0.51	0.70
1pbd		−11.4	−11.4	−9.9	0.32	0.26	3hvt		−11.9	−11.4	−9.0	0.72	0.79
1pgp	−7.8	−6.3	−6.3	−8.8	2.23	1.83	3mth		−5.4	−5.4	−5.9	1.23	5.62
1pha		−12.2	−12.2	−8.8	1.04	0.52	3ptb	−6.1	−6.5	−6.5	−8.8	0.23	0.16
1phd		−7.4	−7.4	−6.5	1.13	0.30	3tpi	−5.9	−9.8	−7.8	−9.0	0.47	0.51
1phf	−6.0	−8.2	−8.1	−6.3	1.46	1.11	4aah		−7.6	−7.6	−11.0	0.25	0.24
1phg	−11.8	−10.9	−10.9	−8.1	4.29	1.20	4cts		−4.6	−4.6	−8.8	0.25	0.19
1poc	−10.7	−8.4	−8.4	−10.1	1.44	1.52	4dfr	−11.8	−10.1	−10.1	−9.3	0.92	5.17
1ppc	−8.4	−11.9	−9.9	−9.8	1.40	6.31	4fab	−11.0	−13.7	−12.9	−9.1	4.44	0.82
1pph	−8.1	−10.5	−8.5	−10.4	0.70	0.58	4fbp		−9.7	−9.7	−12.1	2.03	0.55
1ppi		−16.7	−14.7	−8.5	1.01	2.80	4fxn		−17.3	−17.3	−13.2	0.50	0.49
1ppk	−10.4	−9.9	−9.9	−10.1	0.73	0.27	4hmg	−3.5	−6.6	−6.6	−6.7	0.54	0.67
1ppl	−11.7	−10.3	−10.3	−11.6	0.70	2.55	4phv	−12.5	−14.5	−14.5	−11.8	0.55	4.22
1ppm	−7.9	−12.5	−12.5	−11.7	0.99	0.62	4tim	−2.9	−5.3	−5.3	−10.2	1.32	1.32
1pro	−15.4	−15.6	−15.6	−12.7	1.50	1.51	4tln	−5.1	−7.1	−7.1	−6.6	1.43	2.67
1pso	−14.1	−10.9	−10.9	−9.6	5.15	6.12	4tmn	−13.9	−11.9	−11.9	−12.1	1.29	0.73
1sbg	−10.6	−11.1	−11.1	−10.9	0.88	0.40	4tpi	−4.0	−6.6	−6.6	−8.5	0.88	0.56
1slt		−5.8	−5.8	−6.2	1.03	0.57	4ts1	−6.7	−8.6	−8.6	−8.8	0.89	0.85
1snc	−9.1	−9.6	−9.6	−9.8	2.06	1.12	5abp	−9.1	−7.6	−7.6	−8.5	0.11	0.20
1sre	−5.3	−8.9	−8.5	−10.0	0.30	0.36	5cpp	−8.0	−8.2	−8.2	−6.8	0.11	2.65
1srj		−13.7	−12.9	−10.0	0.47	0.49	5cts		−3.2	−3.2	−8.1	0.27	0.27
1stp	−18.3	−18.2	−18.2	−10.7	0.60	0.58	5p2p		−9.0	−9.0	−5.4	4.95	6.18
1tdb		−7.9	−7.9	−7.6	7.34	7.50	5tim	−3.1	−3.2	−3.2	−7.3	1.32	0.69
1thy		−7.0	−7.0	−7.2	1.98	4.21	5tln	−8.7	−12.4	−12.4	−9.9	2.37	1.01
1tka		−7.7	−7.7	−11.5	2.28	2.28	5tmn	−11.0	−12.0	−12.0	−12.1	2.87	2.50
1tlp	−10.3	−10.0	−10.0	−10.4	7.70	7.33	6abp	−7.7	−7.7	−7.7	−8.8	0.33	0.36
1tmn	−10.0	−10.2	−10.2	−10.4	3.65	1.90	6cpa	−15.7	−10.8	−10.8	−10.8	3.93	4.29
1tng	−4.0	−3.9	−3.9	−8.3	0.26	0.21	6rnt	−3.2	−3.7	−3.7	−8.2	0.63	0.63
1tnh	−4.6	−3.9	−3.9	−8.7	0.39	0.28	6tim	−4.4	−5.9	−5.9	−8.3	0.59	0.42
1tni	−5.5	−3.7	−3.7	−6.7	2.12	2.03	6tmn	−6.9	−10.4	−10.4	−11.6	2.53	2.57
1tnj	−2.7	−4.1	−4.1	−7.8	0.43	0.36	7abp	−8.6	−8.0	−8.0	−9.4	0.15	0.17
1tnk	−2.0	−3.8	−3.8	−6.9	1.17	0.98	7cpa	−19.0	−13.1	−13.1	−10.5	3.91	2.87
1tnl	−2.6	−4.0	−4.0	−7.4	0.54	0.24	7cpp	−5.2	−6.2	−6.2	−5.7	1.99	3.23
1tph	−3.1	−5.3	−5.3	−7.5	0.22	0.23	7tim	−7.4	−5.4	−5.4	−7.7	0.20	0.19
1tpp		−7.1	−5.1	−7.9	0.43	1.07	8abp	−10.7	−7.6	−7.6	−8.4	0.10	0.21
1trk		−9.6	−9.6	−11.2	1.63	2.15	8atc	−10.3	−9.5	−9.5	−10.7	0.41	0.38
1tyl		−5.2	−5.2	−6.7	5.20	1.08	8gch		−7.9	−7.9	−9.6	0.32	0.30
1ukz		−11.3	−11.3	−13.6	0.55	0.41	9abp	−10.9	−7.2	−7.2	−10.4	0.23	0.13
1ulb	−6.0	−6.7	−6.7	−8.7	0.34	0.35	9hvp	−11.4	−11.9	−11.9	−12.9	1.44	1.47

[a] RMS values are computed relative to the native cocrystallized ligand. Scores are compared to available experimental binding energies, ΔG_{exp} (in kcal/mol). Note that the XP scores have been adjusted to cap the hydrophobic enclosure packing term at a value of −2.0 to bring this term closer to an absolute energy scale. Larger RMS values are indicated with a (*) in cases where a nearly chemically symmetric solution was found by XP docking. [b] The pdbbind dataset,[42] v2002, lists a binding affinity of −13.6 kcal/mol, but this is for the entire antigen−antibody complex, whereas the structure provided is for a fragment of the complex that appears to lack essential antigen−antibody interactions.

affinities. This idea could form the basis for an approach enabling scores achieved in different protein conformations for a given receptor to be related based on experimental calibration of reorganization energy. Performance of the methodology when there is substantial reorganization is addressed in a preliminary fashion by the enrichment studies in section 4, where a number of such ligand−receptor pairs are considered. In these cases, knowledge of the reorganization energy of the receptor is not necessary to rank order the binding of compounds to a particular form of the receptor.

To address less extreme yet still nontrivial reorganization effects present in our data set, we define "average" adjustable parameters to convert calculated empirical scores into predicted experimental binding affinities. In particular, we estimate the

Table 6. RMS and Average Absolute Deviations (Avg) in Predicted Binding Affinities for the XP 4.0 and SP 4.0 Scoring Functions[a]

		XP 4.0		SP 4.0	
comparison	number	RMSD	avg	RMSD	avg
all ligands with ΔG_{exp}	198	2.26	1.75	3.18	2.51
all well-docked ligands	136	1.73	1.34	2.80	2.23
all poorly docked ligands	62	3.02	2.49	3.70	3.11

[a] RMS and average absolute deviations are presented in kcal/mol. Well-docked ligands are defined as those having an RMSD to the cocrystallized pose of 2.5 Å or less, plus those identified in Table 5 as being docked appropriately but in a symmetry-related orientation.

Table 7. Training Set Used to Characterize XP Virtual Screening[a]

PDB code	description	no. actives	no. well-docked actives
1e66	acetylcholinesterase	20	20
1bji	neuraminidase	9	9
1fjs	factor Xa	13	8
1kv2	human p38 map kinase	10	10
1bl7	p38 map kinase	36	27
1rt1	HIV-RT	29	23
1cx2	cyclooxygenase-2	13	13
1aq1	human cyclin dep. kinase	10	6
1ett	thrombin	16	15
1hpx	HIV-1 protease	14	9
3ert	human estrogen receptor	10	8
1qpe	lck kinase	121	87
1m17	EGRF tyrosine kinase	117	106
1tmn	thermolysin	6	5
1kim	thymidine kinase	4	4

[a] All correctly docked ligands have experimental activities <10 μM except those for neuraminidase.

hydrophobic-enclosure reorganization energy by capping the maximum assignable enclosure energy at a value of 2.0 kcal/mol, as compared to the maximum scoring-function value of 4.5 kcal/mol. It also appears as though the special bidentate charged−charged reward when the ligand group is positively charged involves some degree of protein reorganization, presumably to enable the bidentate salt bridge to be made properly. Thus, this parameter, which normally is 2.0 kcal/mol, was set to 1.0 kcal/mol for comparison of predicted binding affinities to experiment. Other such average effects could be defined, but due to the limited experimental data considered in this paper, no further attempts were made to determine additional parameters.

Tables 6 and 7 present results for both 4.0 XP and 4.0 SP Glide for structural RMSDs (taken with respect to the refined structure of the native ligand generated by our standard protein-preparation procedure,[1,40] using heavy atoms only) and for binding affinities of the various complexes discussed herein. Table 6 further divides the comparison according to whether the XP complex is roughly correctly docked, that is, has a structural RMSD of 2.5 Å or less. Over all 198 complexes examined, the average RMSD in predicted binding affinity for the 4.0 XP scoring function is 2.3 kcal/mol and the average unsigned error is 1.8 kcal/mol. When only well-docked ligands are examined, the average RMSD for the 4.0 XP scoring function is 1.7 kcal/mol and the average unsigned error is 1.3 kcal/mol. This is a significant improvement over the performance of the 4.0 SP scoring function and is also an improvement over the performances of other empirical scoring functions in the literature of which we are aware. In only 11% (15 of 136) of well-docked ligands were errors greater than 3 kcal/mol, suggesting, at least when the appropriately fitting protein structure is presented to the ligand and the ligand is well-docked, the present scoring function has a respectable ability to distinguish weak (mM), moderate (μM), and strong (nM) binders. This capability is essential to the principal task in virtual screening, yet is only marginally present in prior scoring functions.

To improve beyond this level with regard to precision, we believe that receptor flexibility must be introduced and an additional level of detail with regard to the protein−ligand interactions must be incorporated. Robustness is another matter; new moieties and chemistries seem to emerge as additional receptors are added to the test suite. Thus, we cannot claim to have reached convergence in this regard with our current data sets. On the other hand, the amount of detailed medicinal-chemistry information incorporated into the current scoring function is a substantial advance as compared to alternative scoring functions in the literature.[8−19]

A final important point is that the accuracy cited above may quantitatively degrade in cross-docking calculations, even when the ligand is able to assume a qualitatively correct pose in the receptor, as of course would the accuracy of other docking/scoring methods, for similar reasons. The enrichment studies presented below address this question to some extent, but at present do not consider the relative rankings of different active compounds. We have carried out some preliminary investigations (data not shown) that suggest qualitatively reasonable results can be obtained in a limited number of test cases for ranking active compounds when cross docking is required, but these results are not yet robust across a wide range of receptors, and the quantitative precision with which this can be done is not yet clear.

4. Enrichment Studies

The goal of the present work is to optimize a scoring function that will properly assign binding affinities to active compounds if the compound is docked in a sufficiently native-like pose and that will minimize the number of inactive database ligands that score well. The question of what is sufficiently native-like is a heuristic one, but we would argue that precision in drawing the line is not critical. If a few additional compounds are included, or excluded, this will have minimal effect on the overall optimization process. Therefore, visual inspection has been employed to construct data sets of active compounds that "fit" into the particular versions of the receptors employed. Typically, 60−100% of the available active compounds per receptor fell into this category, so the current protocol is not simply cherry-picking of a small number of compounds. On the other hand, the fact that poorly fitting compounds are not included in the data set must be taken into account when comparing with other results reported in the literature.

We have divided our data set into a training set and a test set. The training set screens have been used in parametrizing the XP scoring function, while the test set screens have not. The training set contains 15 receptors and various numbers of ligands for each receptor, as enumerated in Table 7. A large and diverse training set is essential to address the range of chemical motifs identified in the XP Glide scoring function, as noted previously. Our present test set is relatively small and less diverse than the training set with regard to the number of new receptors considered. Therefore, the validation implied by the results must be considered preliminary. Six receptors are considered in the test set, four of which were not included in the training set: vascular endothelial growth factor receptor 2 (Vegfr2), peroxisome proliferators activated receptor γ (PPARγ), β-secretase (BACE), and blood coagulation factor VIIa (factor VIIa). For two training set receptors, CDK2 and thrombin, we

have located a significant number of additional ligands to include as part of the test set.

Training Set Results. Table 7 lists, for the training set, the receptors investigated, the number of known active ligands available with affinities better than 10 μM, and the number of ligands deemed to dock correctly into the chosen receptor. There is one exception; active neuraminidase ligands are relatively weak binders that do not have activities better than 10 μM. As suggested above, we assume that a random database will contain relatively few compounds with potency greater than this value. Given the process defined above, a key objective of the paper is to demonstrate the improvement that is obtained from employing the new terms defined above. Comparison of version 4.0 XP Glide is made with the following:

(1) Version 4.0 SP Glide, which has been optimized using a similar training set. The 4.0 SP Glide scoring function includes some XP terms, but with very small coefficients. In addition to terms such as those in ChemScore, from which SP Glide was originally derived, this scoring function also includes contributions from the Coulomb and vdW protein−ligand interaction energies and from Schrödinger's "active-site mapping" technology.

(2) A preliminary version of XP Glide (v2.7). This version of XP had a nonoptimal set of penalty terms, an initial version of the hydrophobic-enclosure term, a crude representation of the special hydrogen-bond reward term without the crucial coupling to the hydrophobic-enclosure term, and a sampling methodology significantly inferior to that in Glide XP 4.0. This comparison enables assessment of the impact of the improved sampling and scoring methodologies used in Glide XP 4.0.

Protein and Ligand Preparation. Because the XP Glide scoring function is based on enforcement of physical chemical principles to a much greater degree than is employed in many other scoring functions, appropriate protein and ligand preparation is particularly critical. In practical applications, it is often necessary to carry out such preparation without prior knowledge of the binding mode of the complex. However, for the present purposes, the objective is to optimize and evaluate the scoring function for correctly docked compounds. Therefore, we have endeavored to use all available information in preparing ligand and protein structures.

The most problematic aspect of protein and ligand preparation is the assignment of protonation states of ligand and protein in the protein active site (note that we neutralize ionizable residues distant from the active site to mimic the effects of dielectric screening by solvent and counterions). In the absence of structural data, correctly making such assignments can be a very challenging task. However, when structural data is available, the most likely protonation states of both protein and ligand can usually be deduced from the structure of the complex. In cases for which we do not have a PDB structure, the correct binding mode of the ligand can typically be inferred by analogy. With a binding mode and solution-phase pK_as of the ligand, the correct protonation states can then be assigned. It should be emphasized that the results shown here require accurate protonation-state assignment, and substantial degradation can result from incorrect assignments in unfavorable cases.

A second aspect of protein preparation is relaxation of the receptor structure so that it at least accommodates the native ligand. We employ the standard Schrödinger protein preparation utility[1,40] for this purpose. A related issue is the use of van der Waals scaling of nonpolar ligand and protein atoms to take minor induced-fit effects into account in an approximate fashion. Various scalings have been examined though, with the exception of the human estrogen receptor (3ert), which used a scaling of 0.8 on the ligand and 0.9 on the protein, the "standard" scaling of 0.8 on the ligand and no protein scaling has been applied. Only active ligands that succeed in 4.0 XP docking with the chosen scaling parameters have been retained in our enrichment studies.

Comparison Database. Our methods for generating comparison databases are outlined in ref 1. Molecules are selected from a purchasable-compound library of about one million compounds that have been filtered for predicted pharmacokinetic properties using the QikProp program.[41] Selection protocols are then applied to ensure a distribution of rings, acceptors, donors, molecular weight, and so on in line with averages determined for drug-like molecules.

Computed Binding Affinities of Known Actives for a Wide Range of Targets. As stated above, our expectation is that only a small number of database ligands will be competitive with active compounds whose experimental binding affinities are better than 10 μM. An ensemble of such compounds can, therefore, be used to optimize the scoring function. Because of the wide range of novel terms that have been incorporated, it has been necessary to perform optimizations using a wide variety of receptors and active compounds. The data set used to date is far from complete in covering the range of potential binding motifs, but is significantly larger and more diverse than any previous data set used in the literature for this purpose.

The optimization process consists of adjusting parameters so that as few database ligands as possible achieve better scores than the tight-binding (better than 10 μM) known actives for the term or terms under consideration. In practice, it is not possible to achieve perfect rank ordering in this regard, for the reasons discussed previously. The average error in binding affinity prediction in the PDB data set is about 1.7 kcal/mol for properly docked ligands, with some outliers with errors of more than 3 kcal/mol. These errors may be somewhat larger when cross docking, rather than self-docking, is performed. On the basis of this analysis, some database ligands with experimental binding affinities in the 10−100 μM range are likely to be computed as having low micromolar or even nanomolar binding affinity, while at the same time some of the active compounds will be underpredicted by similar amounts. As a result, overpredicted database ligands can score ahead of underpredicted active compounds. The crucial goal, however, is not to achieve perfection, but rather to eliminate systematic errors that can lead to large numbers of false positives and little enrichment. Neglect of special neutral−neutral hydrogen-bond terms such as those discussed in section 2 is an example of a systematic error of this nature. One would predict that any method that neglects this term should exhibit poor enrichment factors for kinases such as CDK2, where such hydrogen bonds are a crucial component of the molecular-recognition motif. This is because we have found that large hydrophobic compounds with a few strategically placed polar groups can "fit" into the active site and form structures that, based on the usual pair-scoring terms, are highly competitive with the known actives.

Another, and perhaps the ultimate, measure of performance of the scoring function is whether high-ranking database ligands are in fact active. The best way to address this issue is via experimental testing of such database ligands. We are in the process of carrying out such tests.

Training Set Composition. Our training set, primarily focused on pharmaceutical targets of current interest, is presented in Table 7. This suite represents a wide variety of

Table 8. Average Attractive Components of 4.0 XP Score from Eq 3 of Correctly Docked Active Ligands in the Training Set[a]

screen	$\langle E_{phobic\_pair}\rangle$	$\langle E_{hb\_pair}\rangle$	$\langle E_{hb\_nn\_motif}\rangle$	$\langle E_{hb\_cc\_motif}\rangle$	$\langle E_{hyd\_enclosure}\rangle$	$\langle E_{PI}\rangle$	$\langle E_{bind}\rangle$
acetylcholinesterase	−12.1	−0.3	0.0	−1.9	−4.4	−1.1	−20.8
neuraminidase	−3.8	−0.7	−0.4	−3.6	0.0	0.0	−8.6
factor Xa	−8.3	−0.8	−1.8	−1.2	−0.8	−0.8	−14.1
human p38 map kinase	−10.2	−1.6	−0.3	0.0	−2.6	0.0	−14.7
p38 map kinase	−7.4	−1.2	0.0	0.0	−0.9	0.0	−9.5
HIV-RT	−9.1	−1.0	−1.4	0.0	−4.2	0.0	−15.7
cyclooxygenase-2	−8.9	−0.3	0.0	0.0	−3.2	0.0	−12.5
human cyclin dep. kinase	−7.8	−2.1	−3.2	0.0	−1.5	0.0	−14.7
thrombin	−8.4	−1.7	0.0	−3.0	−0.4	0.0	−13.1
HIV-1 protease	−10.2	−2.6	−0.2	0.0	0.0	0.0	−13.1
human estrogen receptor	−10.6	−1.2	−0.5	0.0	−2.0	0.0	−14.3
lck kinase	−8.1	−1.4	−1.3	0.0	0.0	0.0	−10.8
EGRF tyrosine kinase	−6.7	−1.3	−1.5	0.0	0.0	0.0	−9.5
thermolysin	−8.7	−2.3	0.0	−0.3	−0.1	0.0	−11.4
thymidine kinase	−5.5	−2.3	−3.0	0.0	−1.4	0.0	−12.6

[a] E_{phobic_pair} is the pair lipophilic term (eq 1), E_{hb_pair} is the Chemscore-like pair hydrogen-bond term, $E_{hb\_nn\_motif}$ is the term for neutral−neutral hydrogen bonds in a hydrophobically enclosed environment, $E_{hb\_cc\_motif}$ is the term for special charged−charged hydrogen bonds, $E_{hyd\_enclosure}$ is the hydrophobic enclosure reward, E_{PI} is the pi-stacking/pi-cation reward, and E_{bind} is the sum of all terms in eq 3 that account for favorable binding affinities.

different types of active sites and binding motifs. A rough classification of these is as follows:

(1) Small hydrophobic sites: HIV-RT (1rt1) and cyclooxygenase-2 (Cox-2, 1cx2). The HIV-RT NNRTI site is an allosteric pocket that opens to accommodate the ligand; the Cox-2 site does not display as dramatic a structural rearrangement, but is also highly hydrophobic. The dominant terms of the scoring function are the pair hydrophobic term and the hydrophobic-enclosure term. The hydrophobic enclosure is quite large in both cases for known active compounds. In HIV-RT, the active compounds typically make a single hydrogen bond to a backbone carbonyl that receives the special single neutral−neutral hydrogen-bond reward discussed in section 2. In Cox-2, there is also typically a single hydrogen bond, though it does not receive a special reward.

(2) Medium-sized sites making a single special hydrogen bond. This category includes EGFR tyrosine kinase and the 1bl7 form of p38 MAP kinase. This motif is one of the two typical kinase binding motifs, in which there is a special hydrogen-bonding site in the hinge region of the kinase. These systems allow only a single hydrogen bond in this site, typically involving a ring nitrogen atom, whereas other kinases form a correlated pair or triplet of hydrogen bonds, discussed in (3) below.

(3) Sites making correlated hydrogen bonds. This category includes thymidine kinase (TK), CDK2, and lck kinase (LCK). TK and LCK form a pair of correlated hydrogen bonds in the standard kinase hinge region, while CDK2 actives form either a pair or a triplet, and neutral actives binding to aldose reductase form a correlated triplet similar to that in the streptavidin/biotin pair mentioned above. TK is a small, relatively polar site, with the binding driven primarily by the special hydrogen bonding, whereas CDK2 and LCK are medium-sized and more hydrophobic, binding ligands in the 400−500 molecular weight range, although smaller ligands can bind as well.

(4) Large, buried predominantly hydrophobic sites. This category includes the human estrogen receptor and the 1kv2 conformation of p38 MAP kinase. The estrogen receptor binds large, flat steroid-type molecules and exhibits a medium-sized hydrophobic enclosure term, with the binding driven by this term and by the pair hydrophobic score. The 1kv2 active site is created by an allosteric rearrangement of the p38 activation loop. It has a large hydrophobic enclosure term, and many active compounds make one to two hydrogen bonds. The most active compound, ligand 1kv2, also makes a special hydrogen bond in the hinge region, but binding is primarily driven by hydrophobic terms. The p38 pocket in particular is quite large and, in the absence of the hydrophobic enclosure term, will display a significant number of false positives that bind alternative motifs in the cavity in database screening.

(5) Large, open sites with relatively shallow cavities in the active site pocket. This category includes thrombin (1ett), HIV protease (1hpx), and factor Xa (1fjs). Active ligands in this case are invariably large and fill multiple pockets in the active site. HIV protease actives make a substantial number of hydrogen bonds and derive their binding affinity from this and the hydrophobic pair term. Remarkably, there is no contribution from the hydrophobic enclosure. Thrombin ligands typically form a salt bridge with a buried Asp 189 carboxylate in one of the small available pockets and form other hydrogen bonds as well. Hydrophobic enclosure also makes no contribution for thrombin ligands. Factor Xa exhibits the same salt-bridge motif but also has an unusual hydrophobic location in which a ring moiety of the ligand is sandwiched between a number of aromatic rings in a location near the surface of the protein. This pocket provides some hydrophobic enclosure, although not to the same degree as a deeply buried pocket like that in 1kv2, 1cx2, or 1rt1 and can also accommodate pi-cation and pi stacking interactions. Discriminating false positives that display interactions with the protein surface rather than in binding pockets is important for these receptors, particularly factor Xa.

(6) Small hydrophilic sites in which strong electrostatic interactions are important. This category includes the neuramidase (1bji) and GluR2 receptors. In neuramididase, much of the binding affinity derives from salt bridges, and the various ligands serve as important test cases for the rules for charged−charged interactions laid out in section 2. In GluR2, an unusual set of charged ligands in which charge is distributed over a ring system, are buried in the active site. This system was used to calibrate the use of electrostatic terms to turn off buried-charge penalties and to assign an additional contribution to binding affinity in exceptional cases, as discussed in section 2.

(7) The acetylcholinesterase (1e66) receptor was included to incorporate the well-known pi-cation motif of the active compounds in the parameterization process.

Table 8 lists the average contribution of the hydrophobic enclosure and special hydrogen-bonding terms to the scores for known active compounds that bind to each of the above targets, as well as the average total score. Note that the total score cannot be directly translated into the predicted binding affinity in all cases because there are a number of targets where allosteric rearrangement of the binding pocket leads to substantial protein

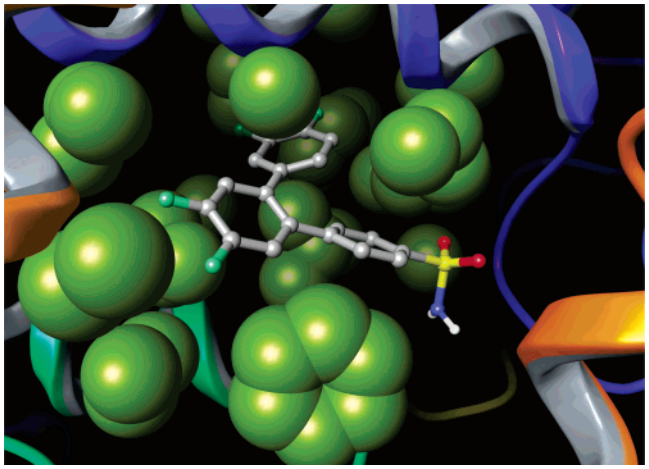

Figure 10. SCX-001 bound to cyclooxygenase-2. The three phenyl rings obtain a 3.7 kcal/mol hydrophobic enclosure packing reward for occupying the large hydrophobic pocket.

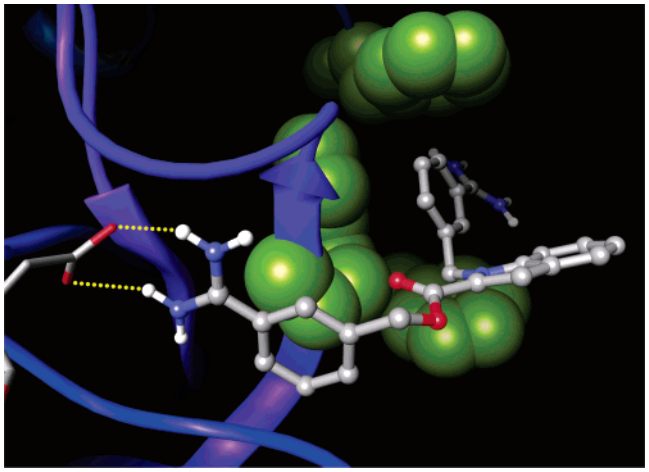

Figure 11. The 1lpk ligand bound to factor Xa. A special salt bridge pair with ASP 189, analogous to that in thrombin (Figure 8), is observed. A hydrophobic enclosure packing reward of −0.8 kcal/mol is achieved by the phenyl ring occupying the Tyr 99/Trp 215/Phe 174 pocket. Also, a pi-cation interaction is received by the charged end of the ligand.

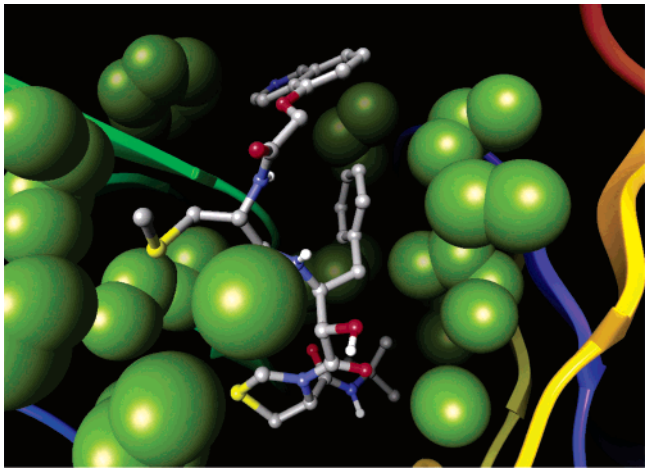

Figure 12. The 1hpx ligand bound to HIV-1 protease. Hydrophobic groups of the ligand are not hydrophobically enclosed and do not receive a hydrophobic enclosure packing reward. For example, the phenyl ring faces hydrophobic residues on only one face.

reorganization energy. Figures 2−13 provide illustrative examples of a "typical" active ligand binding to the various receptors illustrating features that contribute to the specialized scoring-function terms as appropriate.

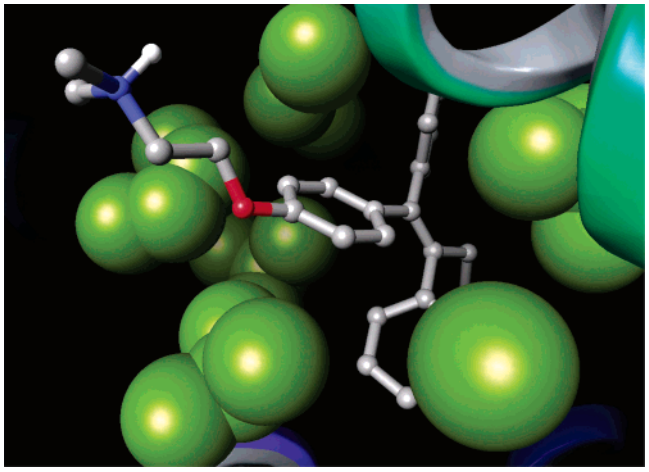

Figure 13. 4-Hydroxytamoxifen bound to the human estrogen receptor. Hydrophobic enclosure about the phenoxy group is illustrated by displaying lipophilic protein atoms as green spheres.

Table 9. Average Enrichments Defined as the Average Number of Outranking Decoy Ligands over Correctly Docked Actives in the Training Set[a]

	avg number of outranking decoys		
screen	v4.0 XP	v2.7 XP	v4.0 SP
acetylcholinesterase	111	580	344
neuramididase	25	411	37
factor Xa	1	196	187
human p38 map kinase	26	30	57
p38 map kinase	7	93	183
HIV-RT	11	26	83
cyclooxygenase-2	24	12	22
human cyclin dep. kinase	3	77	216
thrombin	2	57	70
HIV-1 protease	16	60	167
human estrogen receptor	14	2	23
lck kinase	13		157
EGRF tyrosine kinase	41	411	279
thermolysin	9	107	32
thymidine kinase	0	1	29

[a] Active ligands have at least 10 μM activity except those for neuraminidase, as described in Section 4.

Results of Training Set Enrichment Studies. Table 9 reports a measure of enrichment defined as the average number of database ligands outranking the active compounds in the database. Specifically, the number of database ligands with a GlideScore that is superior to each active is tabulated, these values are summed, and the result is then divided by the total number of active compounds in the data set. We believe that this metric is superior to standard definitions of enrichment, which punish active ligands when they are outranked by other active ligands; this is a particularly serious problem when the active test suite contains a large number of compounds. A "perfect" score based on this metric would thus be zero (no database ligands outranking any active compounds), and smaller numbers are better. These values are also presented for the older 2.7 XP and 4.0 SP Glide results. As noted previously, only active ligands that successfully docked in 4.0 XP Glide were considered. In a small number of cases, active ligands failed to dock with 4.0 SP or 2.7 XP. For a calculation of the number of outranking decoy ligands, such ligands were ranked lower than all successfully docked active and decoy ligands.

Table 10 is the corresponding table constructed using a more standard definition of enrichment that we have employed

Table 10. Standard Enrichment Factors for Recovering 40% of the Correctly Docked Active Ligands in the Training Set[a]

	enrichment factors		
screen	v4.0 XP	v2.7 XP	v4.0 SP
acetylcholinesterase	37	1	2
neuramididase	64	2	112
factor Xa	126	42	63
human p38 map kinase	81	58	51
p38 map kinase	35	18	11
HIV-RT	36	14	18
cyclooxygenase-2	35	56	78
human cyclin dep. kinase	168	168	8
thrombin	68	58	12
HIV-1 protease	90	32	112
human estrogen receptor	126	126	126
lck kinase	12		6
EGRF tyrosine kinase	6	1	2
thermolysin	50	134	201
thymidine kinase	251	251	17

[a] Active ligands have at least 10 μM activity, except those for neuraminidase, as described in section 4.

previously.[2] The results in Table 10 include the same sets of active ligands as in Table 9, that is, those whose activities are better than 10 μM and have been judged to fit more or less correctly into the specified conformation of the receptor. Table 11 presents results using all active compounds with binding affinities better than 10 μM, whether the binding mode is judged to be correct. Results are presented for recovering 40, 70, and 100% of the considered active ligands in each case. This type of analysis corresponds to the approach taken by us in ref 2, as well as to other work in the literature. While we believe that the analysis in Table 9 is the appropriate one to use in assessing the quality of a scoring function to be used in rigid-receptor docking, the results presented in Table 11 enable a direct connection to be made with alternative viewpoints. In what follows, our discussion is focused on the results in Table 9 for the reasons given above.

The XP 4.0 results are nearly uniformly comparable to or better than those of either SP 4.0 or XP 2.7 and, in many cases, are significantly better, as is manifested with particular clarity using the new definition of enrichment. There is a slight degradation for the estrogen receptor from XP 4.0 for cyclooxygenase-2 relative to both XP 2.7 and SP 4.0, but all of the results for these test cases are very good. The real question with regard to scoring-function effectiveness is the ability to prevent false positives from ranking ahead of active compounds. XP 4.0 displays an ability to reduce the average number of false positives ranking ahead of actives in many cases by an order of magnitude and in some cases by nearly 2 orders of magnitude, as compared to both 2.7 XP and 4.0 SP. This same effect is also reflected in the more common definition of enrichment factor (Table 10), but the improvement is quantitatively obscured by the definition of enrichment employed, particularly for the data sets containing larger numbers of actives. For example, in EFGR kinase the number of actives is greater than 10% of the random database, and standard enrichment measures that effectively penalize active compounds for having other active compounds ranking ahead of other actives can yield enrichment factors of at most 10.3 for Table 10 and 9.5 for Table 11.

The results shown in Table 9 are not perfect. However, until intrinsic RMS fluctuations in the scoring function can be reduced from the present average of 1.7 kcal/mol for well-docked ligands, the scoring function seems unlikely to systematically perform significantly better without overfitting. The number of high-scoring database ligands reflected in this table is consistent with the estimated experimental population of low micromolar hits in a 1000 molecule random database of drug-like molecules. The acetylcholinesterase receptor appears to manifest the largest systematic errors. This is likely due to our inability to optimize the pi-cation and pi-stacking scoring function terms with high precision because we lack sufficiently diverse examples manifesting these terms. There also remain some difficulties associated with smaller, highly hydrophobic sites, such as Cox-2 and in medium-sized sites with a single special hydrogen bond, such as EGFR. Overall though, the results are reasonably robust across the entire data set and clearly represent a major advance over the results obtained using 4.0 SP or 2.7 XP. Direct comparisons with other codes would require using the same sets of actives and database ligands. Based on anecdotal reports from various sources and from comparison with published data,[1] Glide SP has generally performed at least as well in enrichment studies as, if not better than, alternatives such as GOLD and FlexX. One would therefore expect 4.0 XP to outperform these methods by a margin similar to that seen in Table 9 for 4.0 SP.

Results of Test Set Enrichment Studies. A summary of key data for our test set, including the receptor crystal structures used, and the number of known and well-docked actives again restricted to ligands with experimental binding affinities better than 10 μM are presented in Table 12. The test set includes two kinases (CDK2, Vegfr2), three proteases (thrombin, BACE, factor VIIa), and one nuclear hormone receptor (PPARγ) and, hence, is reasonably diverse with regard to function; all of the receptors in the test set are drug targets of current or recent interest. There is less diversity with regard to active site size and hydrophobicity than in the training set. As discussed above, validation with a larger test set will be addressed in future publications. All test set calculations were performed with the released versions of Glide 4.0 XP and SP, with no parameter adjustment being made to improve results for any targets.

For two of the receptors (Vegfr2 and PPARγ) we utilize two different forms of the receptor structures. These are highly flexible active sites, and a significant fraction of ligands in both cases can be divided into groups that clearly fit better into one version of the receptor or the other. For PPARγ, for example, one class of ligands requires opening of an allosteric pocket (primarily via motion of a phenylalanine residue), while the second class is smaller and does not protrude into this pocket. Comparing scores of these two ligand classes using a single receptor structure does not make sense. If the pocket is closed, the larger ligands will not fit at all, whereas if the pocket is open, the larger ligands will unfairly score better, as the reorganization energy of the receptor required to engender the needed side chain motion will not have been included. There is no overlap between the ligands associated with the two receptor forms. This partitioning is meant as an introductory exploration of enrichment studies using multiple receptor conformations, a topic we intend to pursue more intensively in the future.

Enrichment metrics to recover well-docked active ligands based on number of outranking decoys and standard enrichment measures (as for the training set) are presented in Tables 13 and 14, respectively. For all known active ligands, standard enrichment measures are presented in Table 15. The same comparison database of 1000 decoy ligands employed in training set enrichment studies has been used. As expected, there is some quantitative degradation of the XP results from the training set, but overall the results are qualitatively comparable to the training set results using the outranking decoy metric (which we have argued is the most meaningful for our purposes), and the improvements as compared to SP Glide are, on average, significant. For PPARγ, both methods do reasonably well; this

Table 11. Standard Enrichment Factors for Recovering 40, 70, and 100% of Known Active Ligands in the Training Set, *Including* Misdocked Cases

	enrichment factors								
	v4.0 XP			v2.7 XP			v4.0 SP		
screen	40%	70%	100%	40%	70%	100%	40%	70%	100%
acetylcholinesterase	37	19	1	1	1	1	2	2	1
neuramididase	64	34	6	2	1	1	112	23	4
factor Xa	78	58	3	10	1	1	22	9	1
human p38 map kinase	81	59	4	58	11	9	51	21	3
p38 map kinase	27	14	1	13	3	1	3	2	2
HIV-RT	27	19	3	12	12	0	11	6	0
cyclooxygenase-2	35	29	7	56	29	0	78	58	9
human cyclin dep. kinase	81	20	2	51	4	2	5	3	2
thrombin	64	64	2	54	7	0	11	10	3
HIV-1 protease	72	32	14	27	16	2	72	60	1
human estrogen receptor	101	101	2	101	88	2	101	17	2
lck kinase	9	8	2				4	3	1
EGRF tyrosine kinase	5	6	1	1	1	1	2	2	1
thermolysin	42	52	23	112	37	2	168	34	7
thymidine kinase	251	251	201	251	188	201	17	24	17

Table 12. Test Set Used to Validate XP Virtual Screening[a]

PDB code	description	no. actives	No. well-docked actives
1m4h	BACE	77	34
1dan	factor VIIa	93	40
1fm6	PPARγ (closed form)	93[b]	32[c]
1fm9	PPARγ (open form)	93[b]	25[c]
1y6b	Vegfr2 (closed form)	111[b]	21[c]
1ywn	Vegfr2 (open form)	111[b]	26[c]
1aq1	human cyclin dep. kinase	253	143
1ett	thrombin	40	15

[a] All correctly docked ligands have experimental activities <10 μM. [b] When multiple forms of a receptor are utilized, the number of actives reported here is all known actives with affinities <10 μM. [c] Ligands are assessed as optimally fitting into either the open or closed form of the receptor. For example, in PPARγ, 61% of known active ligands (57 of the 93) were assessed as fitting into either the open or the closed form of the receptor.

Table 13. Average Enrichments Defined as the Average Number of Outranking Decoy Ligands over Correctly Docked Actives in the Test Set[a]

	avg no. of outranking decoys	
screen	v4.0 XP	v4.0 SP
BACE	35	342
factor VIIa	30	75
PPARγ (closed form)	45	48
PPARγ (open form)	44	76
Vegfr2 (closed form)	52	222
Vegfr2 (open form)	69	310
human cyclin dep. kinase	25	206
thrombin	9	52

[a] Active ligands have at least 10 μM activity.

Table 14. Standard Enrichment Factors for Recovering 40% of the Correctly Docked Active Ligands in the Test Set[a]

	enrichment factors	
screen	v4.0 XP	v4.0 SP
BACE	30	25
factor VIIa	19	15
PPARγ (closed form)	9	19
PPARγ (open form)	14	40
Vegfr2 (closed form)	25	10
Vegfr2 (open form)	11	2
human cyclin dep. kinase	8	5
thrombin	64	25

[a] Active ligands have at least 10 μM activity.

Table 15. Standard Enrichment Factors for Recovering 40 and 70% of Known Active Ligands in the Test Set, *Including* Misdocked Cases

	enrichment factors			
	v4.0 XP		v4.0 SP	
screen	40%	70%	40%	70%
BACE	12	3	11	0
factor VIIa	10	6	7	5
PPARγ (closed form)	5	2	7	4
PPARγ (open form)	3	3	11	7
Vegfr2 (closed form)	2	1	1	1
Vegfr2 (open form)	3	2	1	1
human cyclin dep. kinase	4	3	2	2
thrombin	19	3	11	6

is a case where SP scoring performs unexpectedly well, as opposed to suggesting a particular problem with the XP scoring function.

A number of caveats should be emphasized with regard to these results. The test set is small, and there are almost certainly cases where enrichment performance will not be as good as that indicated in Tables 13 and 14. Furthermore, high enrichment with few false positives can only be expected when the active ligands are properly docked. The fact that a significant fraction of actives are not well-docked, even when two receptor conformations are used, indicates that if a diverse set of active compounds is to be robustly separated from a random database without false positives or false negatives, significant work needs to be done to better treat receptor flexibility and ensure reliable docking accuracy. However, we believe that the attempt in this paper to separate scoring function accuracy, docking accuracy, and reorganization energy effects is an essential starting point if truly robust approaches are to be developed.

5. Conclusions

We have described a novel scoring function and an enhanced sampling algorithm, the combination of which constitutes the Glide 4.0 XP docking methodology. The methodology has been tested with a diverse set of ligands and receptors, and has produced large improvements in binding affinity prediction and database enrichment as compared to other scoring functions within Glide.

The potential for providing physical insight into the origins of enhanced binding affinity is, in our view, as important as quantitative improvement of enrichment factors. Visualization of XP Glide terms, as is presented in the Figures of the present paper, can be utilized by modelers and medicinal chemists in the design of new inhibitors. The success of design efforts along these lines in the context of lead optimization would provide

the most convincing evidence that the underlying model of molecular recognition proposed herein has substantial validity. Our hope is that the present paper will facilitate work along these lines by describing in considerable detail the theory that underlies the XP Glide implementation.

Acknowledgment. We thank Mark Murcko, Bob Pearlstein, and Barry Honig for reading preliminary versions of this manuscript and providing useful feedback. We also thank Mike Campbell for assistance in generating graphics.

Supporting Information Available: Detailed descriptions of the algorithms used in hydrophobic enclosure scoring and in scaling special neutral−neutral hydrogen-bond motif rewards. References to experimental binding affinities for all test and training set ligands are also included. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.

(2) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47*, 1750−1759.

(3) Perola, E.; Walters, W. P.; Charifson, P. S. A Detailed Comparison of Current Docking and Scoring Methods on Systems of Pharmaceutical Relevance. *Proteins* **2004**, *56*, 235−249.

(4) Krovat, E. M.; Steindl, T.; Langer, T. Recent Advances in Docking and Scoring. *Curr. Comput.-Aided Drug Des.* **2005**, *1*, 93−102.

(5) Kontoyianni, M.; McClellan, L. M.; Sokol, G. S. Evaluation of Docking Performance: Comparative Data on Docking Algorithms. *J. Med. Chem.* **2004**, *47*, 558−565.

(6) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. *J. Med. Chem.* **2006**, *49*, 534−553.

(7) Farid, R.; Day, T.; Friesner, R. A.; Pearlstein, R. A. New Insights about HERG Blockade Obtained from Protein Modeling, Potential Energy Mapping, and Docking Studies. *Bioorg. Med. Chem.* **2006**, *14*, 3160−3173.

(8) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolinine, G. V.; Mee, R. P. Empirical Scoring Functions: I. The Development of a Fast Empirical Scoring Function to Estimate the Binding Affinity of Ligands in Receptor Complexes. *J. Comput.-Aid. Mol. Des.* **1997**, *11*, 425−445.

(9) Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, L. J.; Freer, S. T. Molecular Recognition of the Inhibitor AG-1343 by HIV-1 Protease−Conformationally Flexible Docking by Evolutionary Programming. *Chem. Biol.* **1995**, *2*, 317−324.

(10) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olsen, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Comput. Chem.* **1998**, *19*, 1639−1662.

(11) Muegge, I.; Martin, Y. C. A General and Fast Scoring Function for Protein−Ligand Interactions: A Simplified Potential Approach. *J. Med. Chem.* **1999**, *42*, 791−804.

(12) Bohm, H. J. The Development of a Simple Empirical Scoring Function to Estimate the Binding Constant for a Protein Ligand Complex of Known 3-Dimensional Structure. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 243−256.

(13) Bohm, H. J. Prediction of Binding Constants of Protein Ligands: A Fast Method for the Prioritization of Hits Obtained from De Novo Design or 3D Database Search Programs. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 309−323.

(14) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A. A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *J. Mol. Biol.* **1996**, *261*, 470−489.

(15) Jones, G.; Willet, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267*, 727−748.

(16) Ewing, T. J.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK 4.0: Search Strategies for Automated Molecular Docking of Flexible Molecule Databases. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 411−428.

(17) Golke, H.; Hendlich, M.; Kelbe, G. Knowledge-Based Scoring Function to Predict Protein−Ligand Interactions. *J. Mol. Biol.* **2000**, *295*, 337−356.

(18) Wang, R.; Lai, L.; Wang, S. Further Development and Validation of Empirical Scoring Functions for Structure-Based Binding Affinity Prediction. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 11−26.

(19) Wang, R.; Lu, Y.; Wang, S. Comparative Evaluation of 11 Scoring Functions for Molecular Docking. *J. Med. Chem.* **2003**, *46*, 2287−2303.

(20) Wallqvist, A.; Berne, B. J. Computer-Simulation of Hydrophobic Hydration Forces on Stacked Planes at Short-Range. *J. Phys. Chem.* **1995**, *99*, 2893−2899.

(21) Lum, K.; Chandler, D.; Weeks, J. D. Hydrophobicity at Small and Large Length Scales. *J. Phys. Chem. B* **1999**, *103*, 4570−4577.

(22) Nicholls, A.; Sharp, K. A.; Honig, B. Protein Folding and Association−Insights from the Interfacial and Thermodynamic Properties of Hydrocarbons. *Proteins* **1991**, *11*, 281−296.

(23) Zhou, R. H.; Huang, X. H.; Margulis, C. J.; Berne, B. J. Hydrophobic collapse in multidomain protein folding. *Science* **2004**, *305*, 1605−1609. Huang, X. H.; Zhou, R. H.; Berne, B. J. Drying and hydrophobic collapse of paraffin plates. *J. Phys. Chem. B* **2005**, *109*, 3546−3552.

(24) Cheng, Y. K.; Rossky, P. J. Surface Topography Dependence of Biomolecular Hydrophobic Hydration. *Nature* **1998**, *392*, 696−699.

(25) Hummer, G.; Rasaiah, J. C.; Noworyta, J. P. Water Conduction Through the Hydrophobic Channel of a Carbon Nanotube. *Nature* **2001**, *414*, 188−190.

(26) Liu, P.; Huang, X. H.; Zhou, R. H.; Berne, B. J. Observation of a dewetting transition in the collapse of the melittin tetramer. *Nature* **2005**, *437*, 159−162.

(27) Wallqvist, A.; Berne, B. J. Molecular-Dynamics Study of the Dependence of Water Solvation Free-Energy on Solute Curvature and Surface-Area. *J. Phys. Chem.* **1995**, *99*, 2885−2892.

(28) Lee, S. H.; Rossky, P. J. A Comparison of the Structure and Dynamics of Liquid Water at Hydrophobic and Hydrophilic Surfaces−A Molecular-Dynamics Simulation Study. *J. Chem. Phys.* **1994**, *100*, 3334−3345. Lee, C. Y.; McCammon, J. A.; Rossky, P. J. The Structure of Liquid Water at an Extended Hydrophobic Surface. *J. Chem. Phys.* **1984**, *80*, 4448−4455.

(29) Sharp, K. A.; Nicholls, A.; Fine, R. F.; Honig, B. Reconciling the Magnitude of the Microscopic and Macroscopic Hydrophobic Effects. *Science* **1991**, *252*, 106−109.

(30) Regan, J.; Breitfelder, S.; Cirillo, P.; Gilmore, T.; Graham, A. G.; Hickey, E.; Klaus, B.; Madwed, J.; Moriak, M.; Moss, N.; Pargellis, C.; Pay, S.; Proto, A.; Swinamer, A.; Tong, L.; Torcellini, C. Pyrazole Urea-Based Inhibitors of p38 MAP Kinase: From Lead Compound to Clinical Candidate. *J. Med. Chem.* **2002**, *45*, 2994−3008.

(31) Hendsch, Z. S.; Tidor, B. Do Salt Bridges Stabilize Proteins−A Continuum Electrostatic Analysis. *Protein Sci.* **1994**, *3*, 211−226.

(32) Waldburger, C. D.; Schildbach, J. F.; Sauer, R. T. Are Buried Salt Bridges Important for Protein Stability and Conformational Specificity. *Nat. Struct. Biol.* **1995**, *2*, 122−128.

(33) Marqusee, S.; Sauer, R. T. Contributions of a Hydrogen-Bond Salt Bridge Network to the Stability of Secondary and Tertiary Structure in Lambda-Repressor. *Protein Sci.* **1994**, *3*, 2217−2225.

(34) Luo, R.; David, L.; Hung, H.; Devaney, J.; Gilson, M. K. Strength of Solvent-Exposed Salt-Bridges. Strength of Solvent-Exposed Salt-Bridges. *J. Phys. Chem. B* **1999**, *103*, 727−736.

(35) Krammer, A.; Kirchhoff, P. D.; Jiang, X.; Venkatachalam, C. M.; Waldman, M. LigScore: A Novel Scoring Function for Predicting Binding Affinities. *J. Mol. Graphics Modell.* **2005**, *23*, 395−407.

(36) Wei, B. Q. Q.; Baase, W. A.; Weaver, L. H.; Matthews, B. W.; Shoichet, B. K. A Model Binding Site for Testing Scoring Functions in Molecular Docking. *J. Mol. Biol.* **2002**, *322*, 339−355.

(37) Perola, E.; Charifson, P. S. Conformational Analysis of Drug-Like Molecules Bound to Proteins: An Extensive Study of Ligand Reorganization upon Binding. *J. Med. Chem.* **2004**, *47*, 2499−2510.

(38) Halgren, T. A. MMFF VI. MMFF94s Option for Energy Minimization Studies. *J. Comput. Chem.* **1999**, *20*, 720−729.

(39) Cho, A. E.; Guallar, V.; Berne, B. J.; Friesner, R. A. Importance of Accurate Charges in Molecular Docking: Quantum Mechanical/Molecular Mechanical (QM/MM) Approach. *J. Comput. Chem.* **2005**, *26*, 915−931.

(40) *Schrödinger User Manuals*, Glide v3.0; Schrödinger, L.L.C.: New York, NY, 1994.

(41) *QikProp v2.3*; Schrödinger, L.L.C.: New York, NY, 2005.

(42) Wang, R.; Fang, X.; Lu, Y.; Wang, S. The PDBbind Database: Collection of Binding Affinities for Protein−Ligand Complexes with Known Three-Dimensional Structures. *J. Med. Chem.* **2004**, *47*, 2977−2980.

JM051256O

Glide 5.5

Quick Start Guide

Schrödinger Press

June 2009

Contents

Document Conventions

In addition to the use of italics for names of documents, the font conventions that are used in this document are summarized in the table below.

Font	Example	Use
Sans serif	Project Table	Names of GUI features, such as panels, menus, menu items, buttons, and labels
Monospace	`$SCHRODINGER/maestro`	File names, directory names, commands, environment variables, and screen output
Italic	*filename*	Text that the user must replace with a value
Sans serif uppercase	CTRL+H	Keyboard keys

Links to other locations in the current document or to other PDF documents are colored like this: Document Conventions.

In descriptions of command syntax, the following UNIX conventions are used: braces `{ }` enclose a choice of required items, square brackets `[ ]` enclose optional items, and the bar symbol | separates items in a list from which one item must be chosen. Lines of command syntax that wrap should be interpreted as a single command.

File name, path, and environment variable syntax is generally given with the UNIX conventions. To obtain the Windows conventions, replace the forward slash / with the backslash \ in path or directory names, and replace the $ at the beginning of an environment variable with a % at each end. For example, `$SCHRODINGER/maestro` becomes `%SCHRODINGER%\maestro`.

In this document, to *type* text means to type the required text in the specified location, and to *enter* text means to type the required text, then press the ENTER key.

References to literature sources are given in square brackets, like this: [10].

Getting Started

This manual contains tutorials designed to help you quickly become familiar with the functionality of Glide, using the Maestro interface. This chapter contains a brief overview of the software and some setup instructions for the tutorials. The tutorial begins in Chapter 2 with the generation of grids from a prepared protein to represent the receptor for docking. In Chapter 3, a set of ligands is docked and scored, and the receptor and ligand poses are examined in Chapter 4. Protein preparation is not covered in this manual: see the *Protein Preparation Guide* for details of this task.

Panel-specific online help is available for all Glide panels. If you need help with a Glide task, click the Help button or see the *Glide User Manual*.

To complete the exercises, you must have access to an installed version of Maestro 9.0 and Glide 5.5. For installation instructions, see the *Installation Guide*.

Exercises in some chapters produce structure files that are needed in subsequent exercises. To allow you to begin at any exercise you choose, these and other necessary files (ligand files, for example) are included with the Glide distribution.

1.1 About Glide and Maestro

Glide is designed to assist you in high-throughput screening of potential ligands based on binding mode and affinity for a given receptor molecule. You can compare ligand scores with those of other test ligands, or compare ligand geometries with those of a reference ligand. Additionally, you can use Glide to generate one or more plausible binding modes for a newly designed ligand. Once you locate favorable structures or bonding conformations with Glide, you can use Liaison or QSite to obtain binding energies for ligand-receptor pairs.

Protein Preparation is usually required for Glide calculations. It can be performed for most protein and protein-ligand complex PDB structures using the Protein Preparation Wizard panel in Maestro. For detailed information, see the *Protein Preparation Guide*.

Maestro is Schrödinger's powerful, unified, multi-platform graphical user interface (GUI). It is designed to simplify modeling tasks, such as molecule building and data analysis, and also to facilitate the set up and submission of jobs to Schrödinger's computational programs. The main Maestro features include a project-based data management facility, a scripting language for automating large or repetitive tasks, a wide range of useful display options, a comprehen-

sive molecular builder, and surfacing and entry plotting facilities. For more detailed information about the Maestro interface, see the *Maestro Overview*, the Maestro online help, or the *Maestro User Manual*.

Maestro comes with automatic context-sensitive help (Auto-Help), Balloon Help (tooltips), an online help facility, and a user manual. For more information on getting help, see page 35. You can also undo some operations in Maestro. For more information, see page 31 of the *Maestro Overview*.

The **Impact** computational engine is the underlying computational program for Glide. It can perform molecular mechanics calculations, either through the Maestro interface or from the command line. For information on running basic Impact jobs, see the *Impact User Manual* or the *Impact Command Reference Manual*.

1.2 Preparing a Working Directory

Before you begin the tutorial you need to create a working directory to keep all your input and output files, and then make a copy of the tutorial files.

UNIX:

1. Set the `SCHRODINGER` environment variable to the directory in which Maestro and Glide are installed:

   csh/tcsh: `setenv SCHRODINGER` *installation_path*
   sh/bash/ksh: `export SCHRODINGER=`*installation_path*

2. Change to a directory in which you have write permission.

   `cd` *mydir*

3. Create a directory by entering the command:

   `mkdir` *directory-name*

4. Copy the structure files to the `structures` subdirectory (*version* is the 5-digit Glide version number):

   `cp -r $SCHRODINGER/impact-v`*version*`/tutorial/structures .`

   This command creates the subdirectory as well as copies the files.

5. In the working directory, create subdirectories named `glide` and `grids`.

   `cd` *directory-name*
   `mkdir glide grids`

Windows:

1. Open the folder in which you want to create the folder that serves as your working directory.

   The default working directory used by Maestro is your user profile, which is usually set to `C:\Documents and Settings\`*username* on XP and `C:\Users\`*username* on Vista. To open this folder, do the following:

   a. From the Start menu, choose Run.

   b. Enter `%USERPROFILE%` in the Open text box and click OK.

2. Under File and Folder Tasks, click Make a new folder.

   You can also choose File > Folder > New.

3. Enter a name for the folder.

   If you want to create a folder inside this folder, open the folder and repeat steps 2 and 3.

4. Open the folder that contains the tutorial files. This folder is in the Schrödinger software installation, which by default is installed at `C:\Schrodinger2009`.

   a. Open an explorer window.

   b. Navigate to the Schrödinger software installation.

   c. Open the `impact-v`*version* folder (*version* is the 5-digit `Impact` version number), then open the `tutorial` folder inside that folder.

5. Drag the `structures` folder to the folder you created in Step 3.

   You can close the `tutorial` folder now.

6. Create folders named `glide` and `grids` in the working folder.

   You can use Step 2 and Step 3 for this task.

1.3 Starting Maestro and Setting the Working Directory

Once you have created the working directory you can start Maestro, and set the Maestro working directory. By default, Maestro writes job files to its working directory. You can change the default in the Preferences panel. If you have changed the default, you should change it back for this tutorial.

UNIX:

If you have followed the directions in the previous section, the SCHRODINGER environment variable should be already set, and you should be in the working directory. You can then skip the first two steps.

1. Set the SCHRODINGER environment variable to the installation directory:

   csh/tcsh: `setenv SCHRODINGER` *installation_path*

   bash/ksh: `export SCHRODINGER=`*installation_path*

   This environment variable is also required to run Glide jobs.

2. Change to the desired working directory:

   `cd` *directory-name*

3. Enter the following command:

   `$SCHRODINGER/maestro &`

   The Maestro main window is displayed, and the working directory is Maestro's current working directory. If you are using an existing Maestro session, you can change the directory by choosing Change Directory from the Maestro menu, navigating to the appropriate directory and clicking OK.

Windows:

1. Double-click the Maestro icon on the desktop.

   You can also use the Start menu. Maestro is in the Schrödinger submenu.

2. From the Maestro menu, choose Change Directory.

3. Navigate to the working directory and click OK.

Receptor Grid Generation

This chapter contains exercises that demonstrate how to use the Receptor Grid Generation panel to set up and start a grid file calculation job. Grid files represent physical properties of a volume of the receptor (specifically the active site) that are searched when attempting to dock a ligand. You will use the grid files calculated in this chapter to dock ligands in later Glide exercises.

If you have not started Maestro, start it now. Before proceeding with the exercises, change the working directory to the `grids` directory. See Section 1.3 on page 3 for instructions on how to do these tasks. When you have finished the setup, you should see a path in the title bar of the main window that ends in grids.

2.1 Importing the Prepared Structures

The complex for this exercise is actually in two files, one containing the receptor and one containing the ligand.

1. Click the Import structures toolbar button.

   

   The Import panel is displayed.

2. From the Files of type menu, ensure that Maestro is chosen.

3. Click Options.

   The Import Options dialog box opens.

4. Ensure that Import all structures, Replace Workspace, and Fit to screen following import are all selected.

5. From the Include in Workspace option menu, ensure that First Imported Structure is chosen.

6. Click Close in the Import Options dialog box.

7. In the Import panel, navigate to the `structures` directory and select the file `1fjs_prep_recep.mae.gz`.

8. Click Open.

   The prepared protein is displayed in the Workspace. The protein structure is displayed in ribbon representation. The structure includes solvent molecules (glycerine) and ions (Ca^{2+} and Cl^{-}), but does not include the ligand.

9. Next import the ligand structure file by clicking the Import structures toolbar button.

   

   The Import panel is displayed.

10. From the Files of type menu, ensure that Maestro is chosen.
11. Click Options.

    The Import Options dialog box opens.

12. Deselect Replace Workspace and Fit to screen following import.
13. Click Close in the Import Options panel.
14. In the Import panel, navigate to the `structures` directory and select the file `1fjs_prep_lig.mae.gz`.
15. Click Open.

    The prepared ligand is displayed in the Workspace in tube representation.

2.2 Defining the Receptor

The receptor structure used for grid generation is taken from the Workspace, so you need to exclude the ligand atoms from consideration as part of the receptor.

1. From the Applications menu in the main window, choose Glide > Receptor Grid Generation.

   The Receptor Grid Generation panel opens with the Receptor tab displayed.

2. In the Define receptor section, ensure that Pick to identify ligand and Show markers are selected, and that, in the option menu, Molecule is chosen.
3. In the Workspace, pick an atom in the ligand molecule.

   Dark green markers appear on the ligand.

4. In the Van der Waals radii scaling section, ensure that Scaling factor is set to the default value of 1.00 (no scaling.)

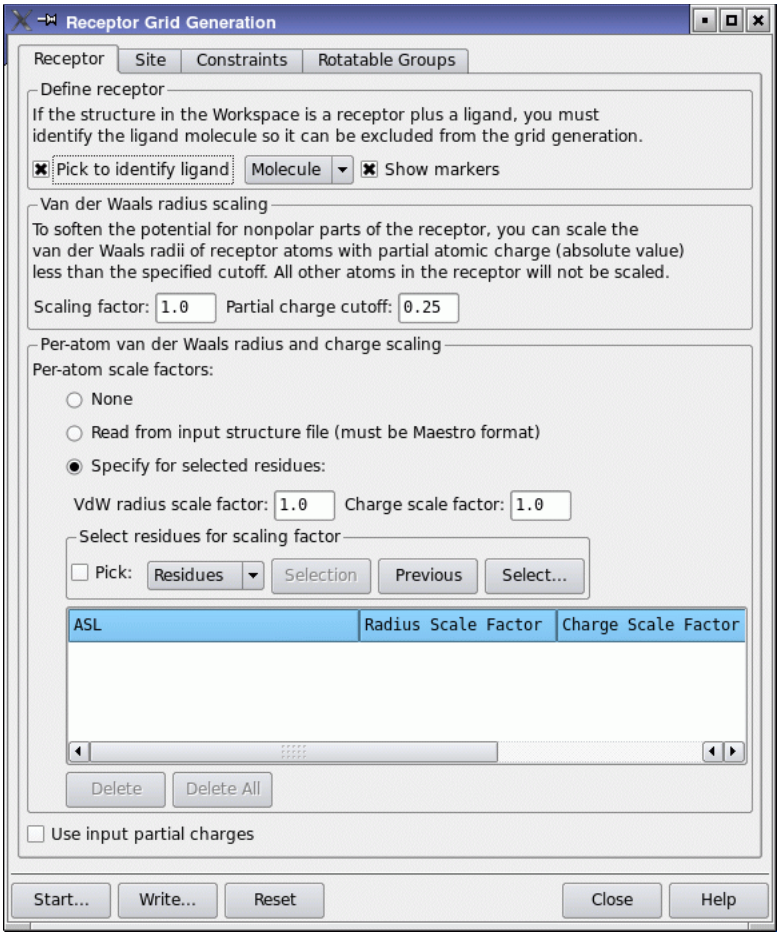

Figure 2.1. The Receptor Grid Generation panel.

2.3 Defining the Active Site

Now that the ligand has been excluded, the volume for which grids will be calculated can be defined:

1. Click the Site tab.

   The entire complex is shown with several types of markers: the dark green ligand molecule markers that appeared when the ligand was identified, and the new markers that appeared when the Site tab was opened:

   - The *enclosing box* is shown in purple.
   - The center of the enclosing box is marked by green coordinate axes.

The purple enclosing box represents the volume of the protein for which grids will be calculated. Generally, you should make the enclosing box as small as is consistent with the shape and character of the protein's active site and with the ligands you expect to dock.

2. In the Site tab, ensure that the Center option selected is Centroid of Workspace ligand.
3. Ensure that the Size option selected is the default, Dock ligands similar in size to the Workspace ligand.

   If you have a representative ligand in the active site, the default generates an enclosing box that is large enough for most systems. However, if you think that conformations of active ligands may exist that are significantly larger than the cocrystallized ligand, you should consider enlarging the enclosing box using the Dock ligands with length <= option.

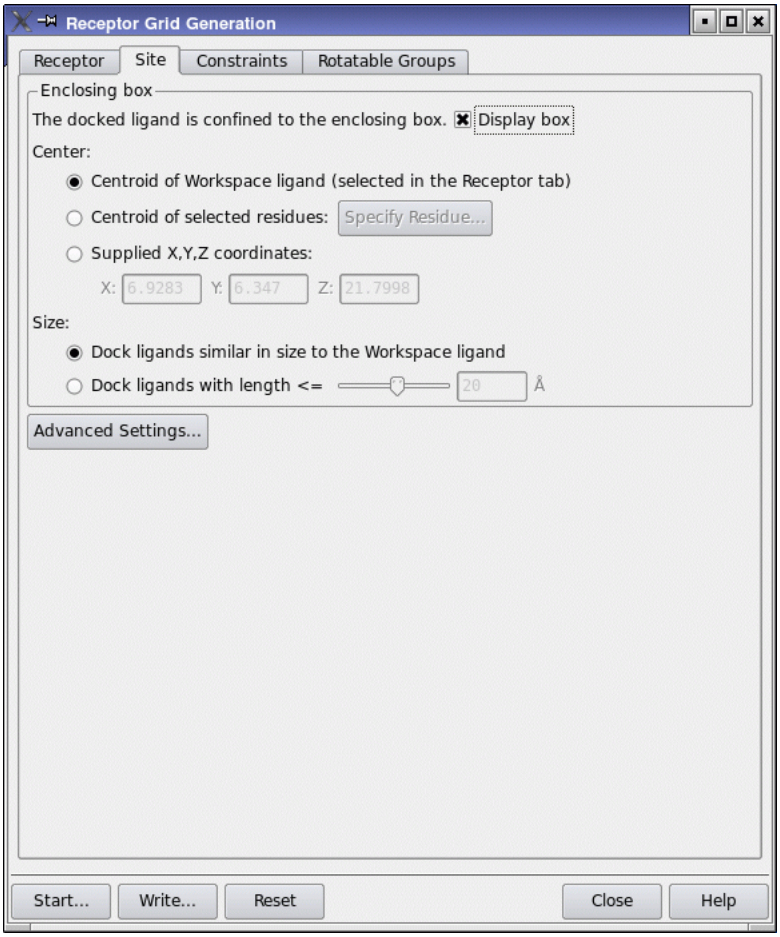

Figure 2.2. The Site tab of the Receptor Grid Generation panel.

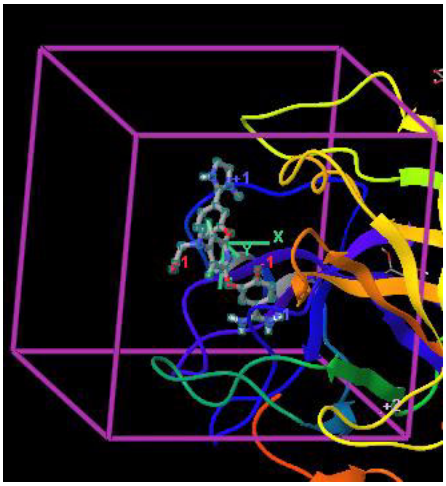

Figure 2.3. The marked ligand with enclosing box.

2.4 Setting Up Glide Constraints

The Constraints tab of the Receptor Grid Generation panel is used to define Glide constraints. In this exercise, you will define two constraints: a positional constraint and an H-bond constraint. To make it easier to see the parts of the receptor close to the ligand and to see the ligand atoms, you will first change the display.

For more information on using Glide constraints, see Section 4.4 of the *Glide User Manual*.

2.4.1 Setting the Display for Constraint Definition

1. From the Display only selected atoms toolbar button menu, choose Molecules, and click on a ligand atom.

   

   The ligand is displayed, and the receptor remains displayed as ribbons.

2. From the Show, hide, or color ribbons button menu, choose Delete Ribbons.

   

3. From Display residues within N Å of currently displayed atoms, choose 3 Å.

   

The residues that are closest to the ligand are displayed.

4. From the Undisplay toolbar button menu, choose Nonpolar hydrogens.

This action leaves the polar hydrogens displayed, making it easier to see the polar hydrogens that are likely to form hydrogen bonds.

2.4.2 Defining a Positional Constraint

1. In the Constraints tab of the Receptor Grid Generation panel, click the Positional tab.
2. Click New.

   The New Position dialog box opens.
3. In the Select atoms to define a position section, ensure that Pick is selected; and from the Pick option menu, that Atoms is selected.
4. Click the carbon atom that is between the two nitrogen atoms in the imidazole ring in the ligand.

   A semi-transparent gray sphere is displayed around the atom.
5. Enter the name `S4_arom` in the Name text box.
6. Enter `2.0` in the Radius text box.
7. Click OK.

   The constraint is added to the Positions table in the Positional tab, and the sphere changes to yellow. The name is displayed next to the sphere.

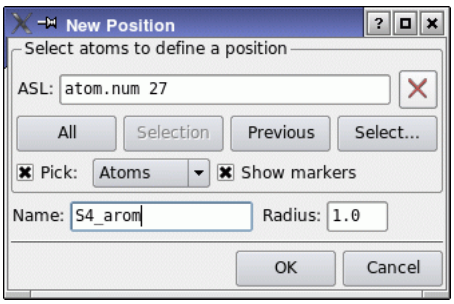

Figure 2.4. The New Position dialog box.

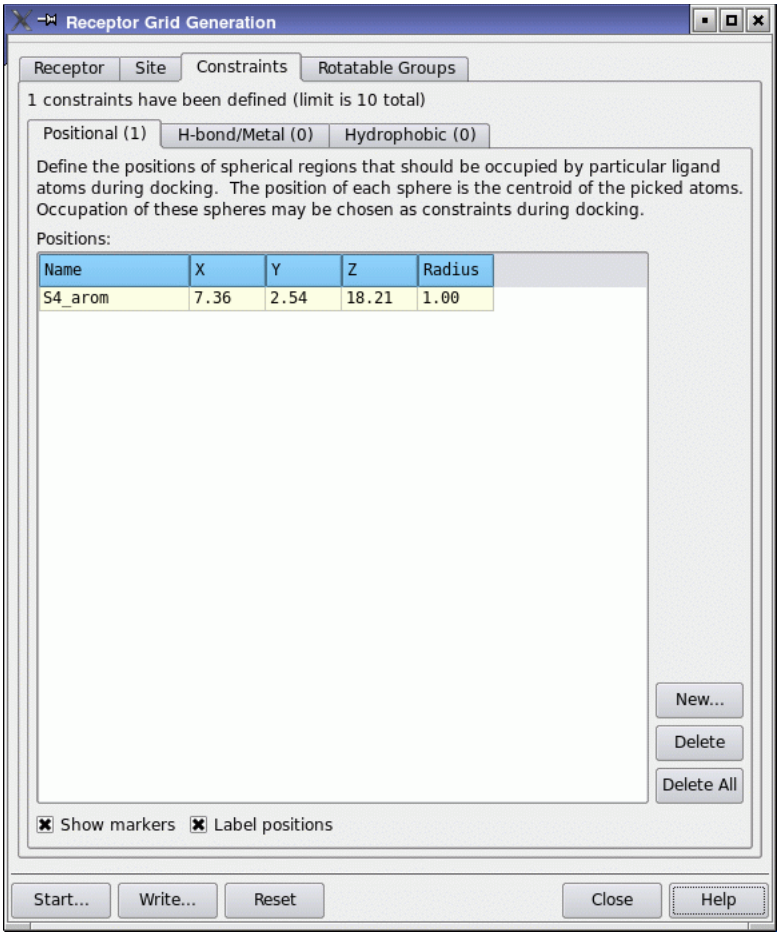

Figure 2.5. The Constraints tab of the Receptor Grid Generation panel showing the Positional subtab.

2.4.3 Defining an H-bond Constraint

Next, you will define an H-bond constraint, for the carboxylate that is H-bonded to the amidine of the ligand. To aid the picking of the constraint, H-bonds to the ligand will be displayed.

1. From the Display H-bonds button menu, choose Inter H-bonds, and click on a ligand atom.

   

   The hydrogen bonds between the ligand and the receptor appear as yellow dashed lines.

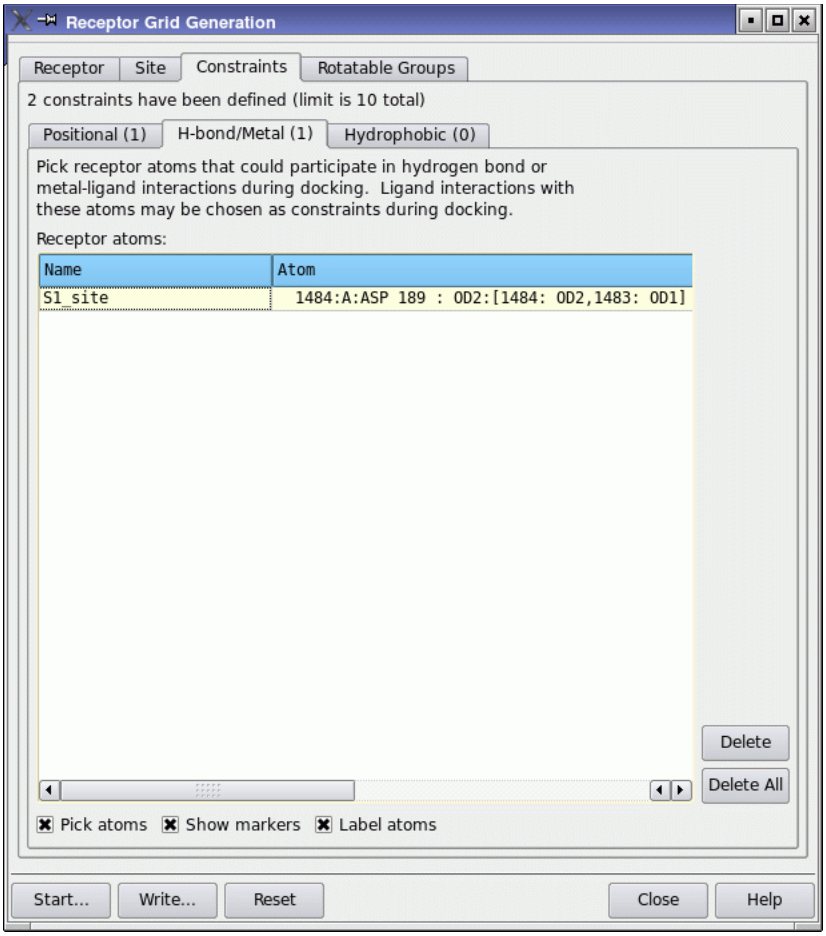

Figure 2.6. The Constraints tab of the Receptor Grid Generation panel showing the H-bond/Metal subtab.

2. In the Constraints tab, click on the H-bond/Metal tab.
3. Ensure that Pick atoms is selected.
4. Click on the carboxyl oxygen that is hydrogen-bonded to the amidine of the ligand.

   This atom is the OD1 atom of ASP 189. When you have picked the atom, an entry is added to the Receptor atoms table. The Name column shows the identity of the atom, with hbond in parentheses. In the Atom column, both oxygen atoms of the carboxylate are listed in square brackets, because Glide includes symmetry-related atoms as part of the constraint. Both atoms are marked in the Workspace with a padlock icon, if Show markers is selected, and the name is also displayed if Label atoms is selected.

5. Name the constraint `S1_site`, by editing the text in the Name column.
6. From the Display H-bonds button menu, choose Delete H-bonds.

2.5 Starting and Monitoring the Grid Calculation

With the ligand and the active site defined, and constraints set up, the grid generation job can be started.

1. In the Receptor Grid Generation panel, click Start.

   You might see a dialog box that warns about an unidentified ligand. This is a check for ligand-sized molecules that might be in the active site. It has found the glycerine molecules, which are not a problem for docking. You can therefore ignore the warning and click Continue.

   The Receptor Grid Generation - Start dialog box is displayed.

2. In the Output section, ensure that the Directory for grid files is the default, `./` , your current working directory, and ensure that Compress is selected.

   When this option is selected, the grid files are placed in an archive that is compressed, so it is easy to copy. Glide can make use of this compressed archive directly, and extracts the files from it as needed.

3. Check the main window title bar to confirm that the current working directory is *yourpath*`/tutorial/grids`.

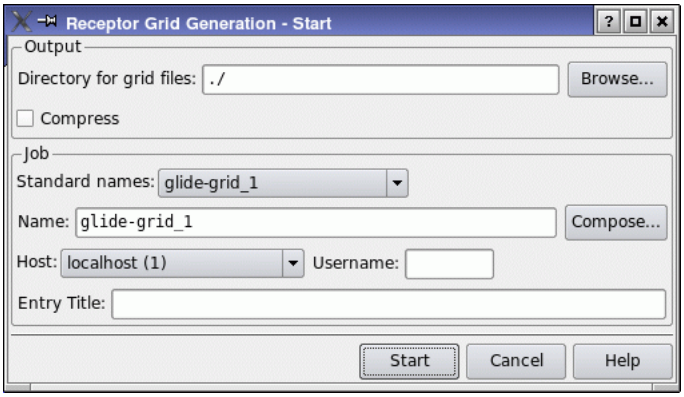

Figure 2.7. The Receptor Grid Generation - Start dialog box.

4. In the Job section, change the Name to `factorXa_grid`.
5. Choose a host and, if necessary, specify a user name.
6. Start the job by clicking Start.

   A warning dialog box might be displayed, informing you that the structure has not been prepared with the Protein Preparation Wizard. You can ignore this warning and click Continue.

   After a moment, the Monitor panel is displayed, and the job starts. While the job is in progress, the Status column displays "running." When the job is complete, the status is changed to "completed : finished". (The opening of the Monitor panel depends on a preference that is set in the Jobs tab of the Preferences panel.)

The job takes approximately 10 minutes on a 1 GHz Pentium 4 processor; this time may vary depending on your particular system configuration and workload.

Before the job is launched, these job input files are written:

`factorXa_grid.in`	Command input for grids job
`factorXa_grid.maegz`	Receptor structure input for grids job

When the calculation is complete, the `grids` directory will contain the following output files:

`factorXa_grid.log`	Log summary file from grids job
`factorXa_grid.out`	Output summary file from grids job
`factorXa_grid.zip`	Archive containing grid files

The zip archive contains the following grid files:

`factorXa_grid.csc`	`factorXa_grid.gsc`
`factorXa_grid.site`	`factorXa_grid.save`
`factorXa_grid_greedy.save`	`factorXa_grid.cons`
`factorXa_grid_recep.mae`	`factorXa_grid.grd`
`factorXa_grid_coul2.fld`	`factorXa_grid_vdw.fld`
`factorXa_grid.vdwc`	

In the next chapter, you will dock ligands using these grids. You can close the Receptor Grid Generation panel now.

Ligand Docking

The exercises in this chapter demonstrate the use of Glide to screen a multiple-ligand file for structures that interact favorably with a receptor active site. The receptor grid files you calculated in the previous chapter will be used to dock ligands from the file `50ligs.mae`. Most of the 50 ligands in the file are decoys, selected as a representative sample from a database of druglike molecules using the `ligparse` utility. Four ligands out of the total of 50 are active ligands for the chosen receptor. These ligands have all been prepared for docking with LigPrep—see Section 3.4 of the *Glide User Manual* or the *LigPrep User Manual* for more information on ligand preparation.

Typically, Glide standard-precision docking is used to find probable good binders in a large set; the top-scoring 10% to 30% can then be investigated more intensively using Glide extra-precision (XP) docking or other methods available from Schrödinger. In these exercises, you will use all three docking modes (HTVS, SP, and XP), and also investigate the use of constraints.

If you have not started Maestro, start it now (see Section 1.3).

Before proceeding with the exercises, change the working directory to the `glide` directory. See Section 2.1 on page 5 for instructions on how to do this.

3.1 Specifying a Set of Grid Files and Basic Options

In this exercise, you will select the grid that you calculated in Chapter 2 for the ligand docking job, and set the basic docking options.

1. Click the Clear workspace toolbar button.

   

2. From the Applications menu, choose Glide > Ligand Docking.

   The Ligand Docking panel opens with the Settings tab displayed.

3. In the Receptor grid section, click the Browse button.

   A file selector opens.

4. Navigate to the `tutorial/grids` directory, choose `factorXa_grid.zip`, and click Open.

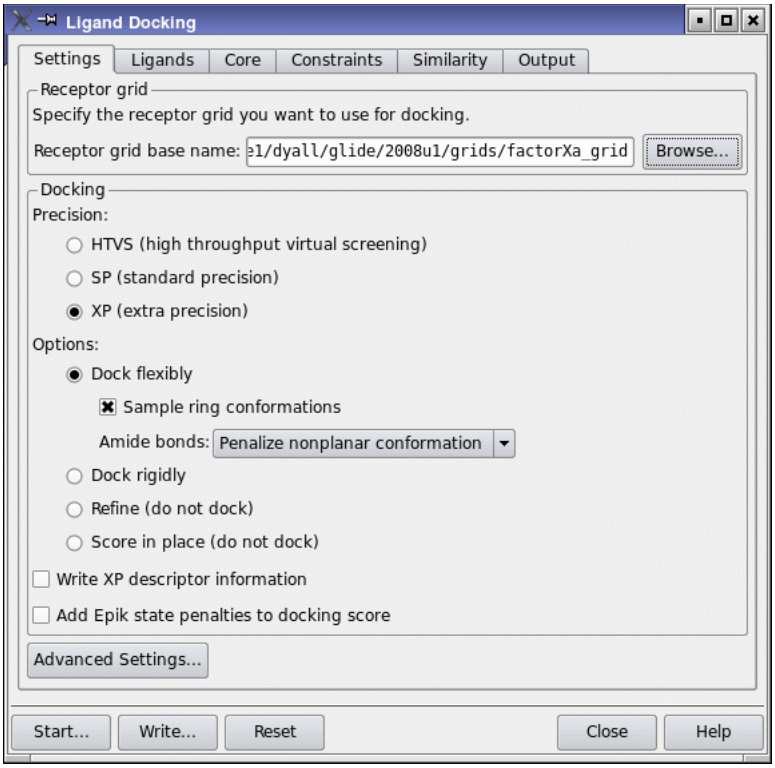

Figure 3.1. The Settings tab of the Ligand Docking panel.

The Receptor grid base name is *fullpath*`/tutorial/grids/factorXa_grid`.

5. In the Docking section, ensure that the Precision option is SP (standard precision).

   This is usually the best choice for docking large numbers of ligands. For more rapid screening you can use the HTVS (high throughput virtual screening) option. You will do this in a later exercise.

6. Under Options, ensure that Dock flexibly and Sample Ring Conformations are selected, and Penalize nonplanar conformation is chosen from the Amide bonds option menu. (These are the defaults.)

The receptor grids and the basic Glide settings for the ligand docking job are now specified. In the next section, you will specify a set of ligands to dock, and in the following section, you will specify output options. The options in the remaining three tabs, Core, Constraints, and Similarity, can be left at their defaults for this exercise.

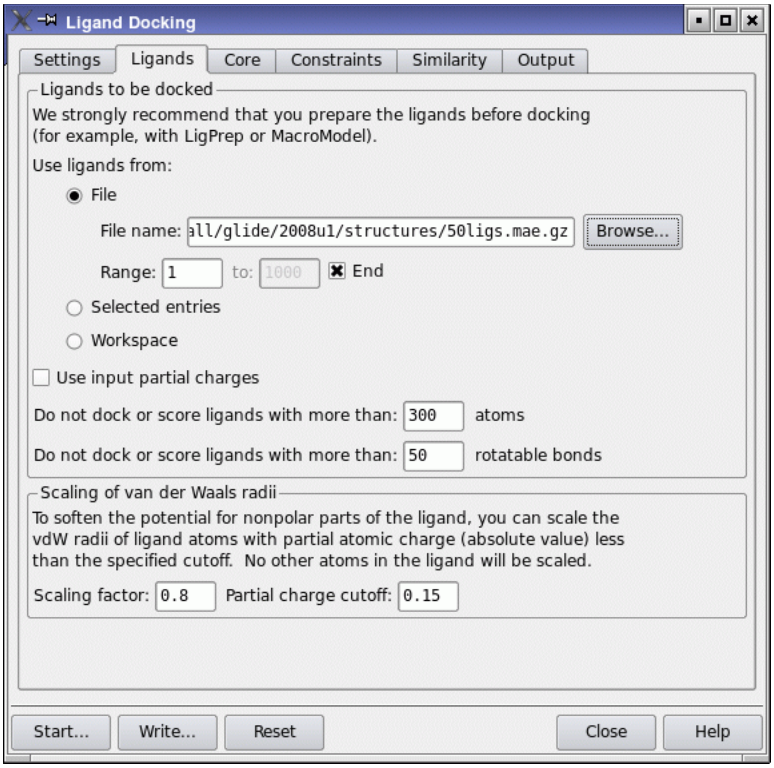

Figure 3.2. The Ligands tab of the Ligand Docking panel.

3.2 Specifying Ligands To Dock

There are several methods for specifying ligand structures to be docked with receptor grids. In this tutorial, you will specify a file containing a set of 50 ligands.

1. In the Ligands tab, ensure that File is selected.
2. Click Browse.

   A file selector is displayed. Ensure that Files of type is set to Maestro
3. Navigate to the `tutorial/structures` directory, choose `50ligs.mae.gz`, and click Open.
4. Ensure that the selected Range is from `1` to End (the default).
5. Ensure that van der Waals radii scaling for ligand atoms is set to the default values: Scaling factor to 0.80 and Partial charge cutoff to 0.15.

In this docking job the constraints that were specified in the grid generation will not be used. In a later exercise you will use the constraints to dock the same ligands.

3.3 Specifying Output Quantity and File Type

The Output tab allows you to specify the type of file to create for the output ligand poses and to determine how many poses to write, per ligand and per docking job.

1. In the Output tab, ensure that Write pose viewer file (includes receptor; filename will be <jobname>_pv.mae) is selected.

   You are specifying that the structural output from the docking job be written to a "pose viewer" file, a file of ligand poses that begins with the structure of the receptor. Having the receptor structure included in the file is convenient for displaying hydrogen bonds and contacts between the ligand and the receptor. In Chapter 4, you will use tools in the Project Table to examine the poses in the file `factorXa_sp_pv.maegz`.

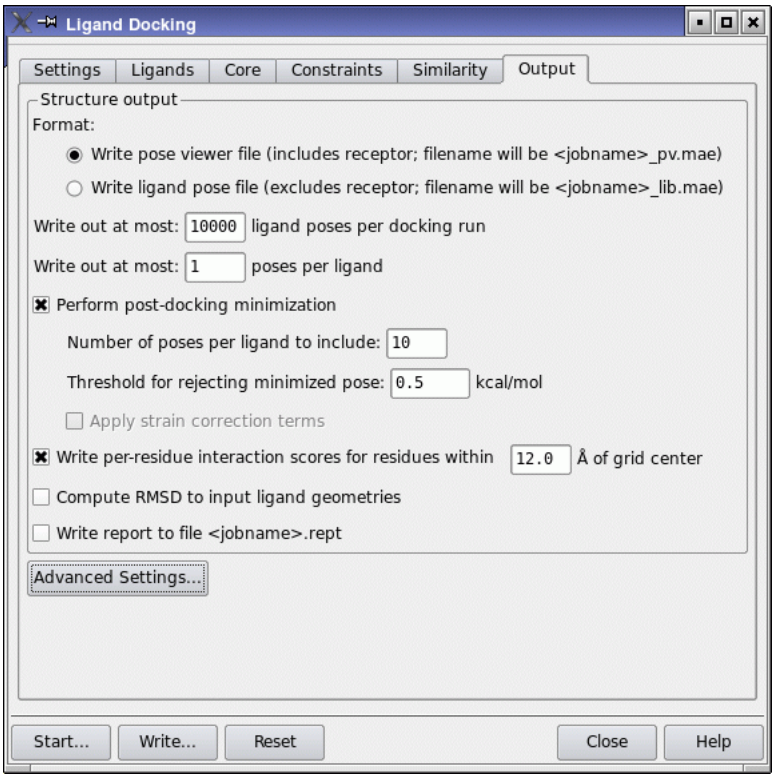

Figure 3.3. The Output tab of the Ligand Docking panel.

2. Ensure that the value of *m* in the Write out at most *m* poses per ligand text box is 1, the default.

   Because there are only 50 ligands in the input file, this setting ensures that no more than 50 poses, one for each ligand, will be collected and written to the pose viewer file.

3. Ensure that Perform post-docking minimization is selected, with the default number of poses per ligand (which is 5).

   Post-docking minimization in the field of the receptor produces better poses and only adds a small amount to the time taken.

3.4 Setting Up Distributed Processing

If you have access to a host machine with multiple CPUs, Glide can divide your multiple-ligand docking job into subjobs that can be distributed over several processors. The Start dialog box allows you to specify the number of subjobs and the number of processors to use. In this section, it is assumed that a host with five or more processors is available. If you have access to such a host, you can follow the instructions in this section without alteration. If the host has fewer than five processors, or you do not want to use five of its processors for this job, change the settings below accordingly.

If you cannot, or do not want to, distribute processing over multiple CPUs, skip to Section 3.5.

1. Click Start.

   The Start dialog box opens.

2. Change the value in the Separate docking into *N* subjobs text box to 10.

   The docking job will be split into 10 subjobs, each one docking 5 ligands. For the sake of speed, the number of ligands per subjob in this exercise is much smaller than would be typical in actual use. The default for *N* is 1, meaning that the job is run as a single job.

3. Choose a multiprocessor host from the Host list table.

   The number of processors on each host, as given in the hosts file, is shown in the Processors column.

4. Edit the value in the Use column for the selected host to specify the number of processors to use.

   The value in the Total to use text box is updated to reflect the value you entered. If you want to use multiple hosts, you can select multiple table rows, and specify the number of processors to use on each. The total number is given in the Total to use text box.

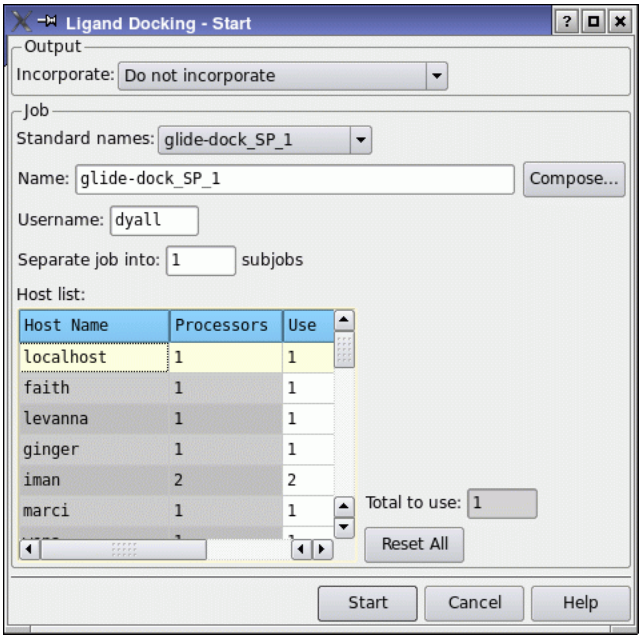

Figure 3.4. The Start dialog box, default settings.

The 10 subjobs will be distributed over the number of processors you specify. We suggest you use five processors. Then the first five subjobs will be sent first, followed by each of the remaining five as processors become available.

When you are satisfied with your distributed processing settings, continue to Section 3.5.

3.5 Starting the Ligand Docking Job

1. If you did not do so in the previous section, click Start.

   The Start dialog box opens.

2. In the Start dialog box, change the name to `factorXa_sp`.
3. From the Incorporate option menu, choose Append new entries as a new group.
4. If you have not already done so, select a host from the Host list.
5. If you have a different user name on the host you have selected from that on your local host, enter the correct user name in the Username text box.
6. Click Start.

   The docking job starts and the Monitor panel is displayed.

For the distributed processing example, as soon as the `factorXa_sp` job has been launched, it is divided into subjobs. As each subjob is launched on a processor, it is listed in the Monitor panel. When one subjob is finished, the next one is launched. To view the log for any subjob, select it in the job table and click Monitor. If the subjob is already finished, the entire log can be scrolled through in the File tab of the Monitor panel. The results for each subjob are stored in subdirectories of the output directory, and collected at the end into the output directory.

The time required for Glide docking jobs depends on the processor speed and workload, the size and flexibility of the ligands, and the volume specified by the enclosing box. As a rough estimate, docking a typical drug-like ligand takes about one minute on a 1.2 to 1.5 GHz Pentium 4 processor under Linux, using Glide 5.5 default settings. On a similar machine with a single processor, the 50-ligand docking job will usually finish in about 45 minutes. As 10 subjobs distributed over 5 similar processors, the docking job will finish in about 10 minutes.

When the job finishes, examine the results in the Project Table panel. The four active ligands (glide lignum 1 through 4) are ranked highest.

3.6 Docking in High-Throughput Virtual Screening Mode

Glide has a set of predetermined options that can speed up the docking by a factor of about seven over the standard precision (SP) docking mode. In this exercise, you will run an HTVS docking job on the same set of ligands as used in the SP docking exercise.

1. In the Settings tab, select HTVS (high throughput virtual screening).

   The other settings will be left as they were for the SP docking job.

2. In the Output tab, ensure that Write ligand pose file (excludes receptor; filename will be <jobname>_lib.mae) is selected.

3. Click Start.

   The Start dialog box opens. You should not need to run this job on multiple processors.

4. Change the job name to `factorXa_htvs`.

5. Select a host, and set the number of processors and subjobs to 1.

6. Click Start.

   The job should take only a few minutes to run.

When the job finishes, examine the results in the Project Table panel. The four active ligands are ranked in the top 5 ligands. Their scores differ a little from those in the SP docking run.

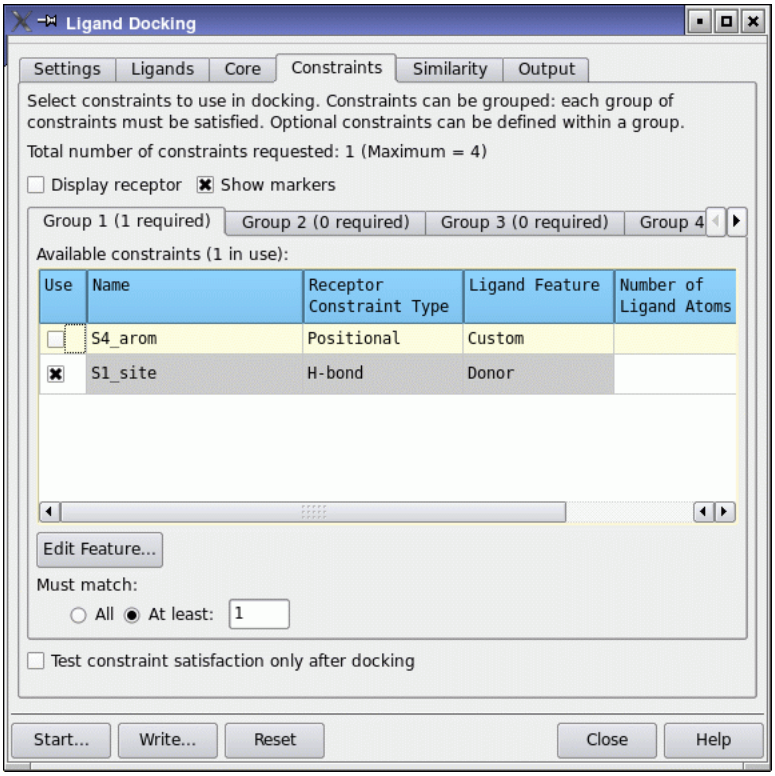

Figure 3.5. The Constraints tab of the Ligand Docking panel.

3.7 Docking Ligands Using Constraints

In this exercise, you will apply the constraints you defined in the grid generation to the docking of the same set of ligands as for the standard SP job. By default, no constraints are applied, even if they are defined. Here, you will require any one of the three constraints to be applied.

1. In the Settings tab, select SP (standard precision).

   The other settings will be left as they were for the HTVS docking job.

2. In the Constraints tab, click both check boxes in the Use column.

   These check boxes mark the constraint for use in docking.

3. Under Must match, select All.

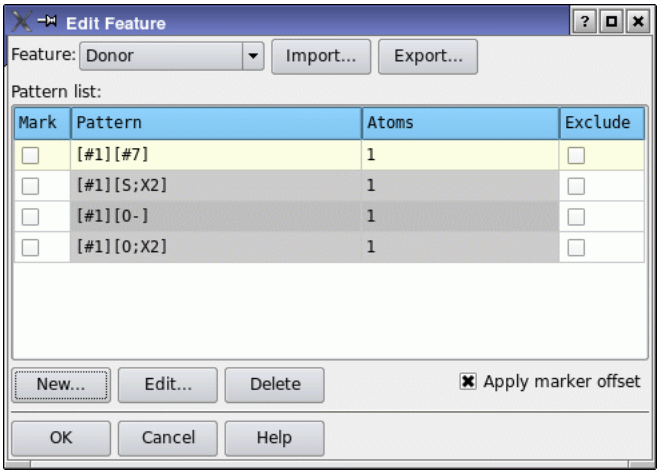

Figure 3.6. The Edit Feature dialog box.

For H-bond and hydrophobic constraints, the ligand features that must match these constraints are predefined. You can edit them if you want, but this is not necessary. For positional constraints, you must define the ligand feature that matches the constraint. Features are defined in terms of SMARTS patterns.

4. From the Available constraints table, select the S4_arom row and click Edit Feature.

   The Edit Feature dialog box opens. There are no SMARTS patterns in the Pattern list table, because the Custom feature type is undefined by default.

5. Click New.

   The New Pattern dialog box opens.

6. Enter the following text into the SMARTS pattern text box:

   `[a]`

   This pattern matches all aromatic atoms.

7. Enter `1` into the Numbers text box.

   This is the index of the atom in the SMARTS pattern that is matched by the constraint. Since there is only one atom in the pattern, this is the only possible choice.

8. Click OK.

   The New Pattern dialog box closes, and a row is added to the Pattern list table in the Edit Feature dialog box.

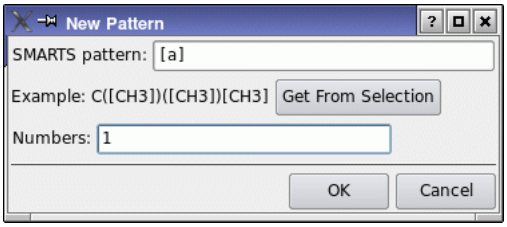

Figure 3.7. The New Pattern dialog box.

9. Click OK.

   The Edit Feature dialog box closes. This completes the definition of the Custom feature. If you did not define this feature, the docking job would not be started.

10. Click Start.

    The Start dialog box opens. If you ran the SP docking job on multiple processors before, you can do the same for this job.

11. Select the host and number of processors.

12. Change the job name to `factorXa_sp_cons` and click Start.

    The Monitor panel opens and displays the progress of the job.

When the job finishes, examine the results in the Project Table. Of the 50 ligands, poses are reported for only six ligands, of which the first four are the actives, and the other two did not score very well. The remaining ligands did not satisfy the constraints. These results indicate that the application of constraints serves to discriminate between ligands that bind in the proper mode, and ligands that don't.

3.8 Docking Ligands Using Core Constraints

In this exercise, you will apply core constraints defined by a pattern in the reference ligand, and use these to dock a set of structures. Structures that do not include the core pattern will not be docked. The first task is to import the reference ligand.

1. In the main window, on the main toolbar, click the Import structures button.

   

   The Import panel is displayed.

2. From the Files of type menu, ensure that Maestro is chosen.

3. Click Options.

   The Import Options dialog box opens.

4. Ensure that Import all structures, Replace Workspace, and Fit to screen following import are all selected.

5. From the Include in Workspace option menu, ensure that First Imported Structure is chosen.

6. Click Close in the Import Options dialog box.

7. Navigate to the `structures` subdirectory and select the file `sar_reference.mae.gz`.

8. Click Open.

   The reference ligand is displayed in the Workspace.

Next, the settings for the previous constraints job need to be cleared.

9. In the Ligand Docking panel, click the Settings tab, and select SP (standard precision).

   The other settings will be left as they were for the previous docking job.

10. In the Constraints tab, clear all check boxes in the Use column.

    Receptor-based constraints will no longer be applied; instead you will be using ligand-based constraints.

The core constraint is defined in the following steps.

11. In the Core tab, select Restrict docking to reference position.

    The first few controls in the Define core section become available.

12. Enter `1.5` in the Tolerance text box.

13. Click an atom in the reference ligand (in the Workspace).

    The ligand is marked with purple markers, and the remaining controls in the Define core section become available.

14. Under Core atoms, select SMARTS pattern.

15. In the main window, from the Undisplay toolbar button menu, choose Nonpolar hydrogens.

    

16. Rotate the structure so that you can clearly see the three six-membered rings.

17. Ensure that the Workspace selection button is selected (indented) and displays an A (for picking atoms).

18. Select the three six-membered rings with their ether linkages, and the amidine group on the terminal ring.

    Do not include the carboxyl on the middle ring or the hydroxyl on the terminal ring.

    You can drag to make the first selection, then hold down the SHIFT key and drag or click to add atoms to the selection. The selected atoms are marked in yellow rather than in purple.

19. In the Ligand Docking panel, click Get From Selection.

    The Smarts pattern text box is filled in with the pattern that corresponds to the atoms selected in the Workspace. The markers on the Workspace selection turn green.

Finally, the ligands to be docked need to be selected. A different set is used from the set used for the previous runs.

20. In the Ligands tab, ensure that File is selected.
21. Click Browse.

    A file selector is displayed. Ensure that Files of type is set to Maestro.

22. Navigate to the `tutorial/structures` directory, choose `sar_series.mae.gz`, and click Open.
23. Ensure that the selected Range is from `1` to End (the default).
24. Start the job with the name `factorXa_sp_core`.

    The job takes a similar time to the constraints job.

25. When the job finishes, examine the results in the Project Table.

3.9 Refining Docked Ligands with Glide XP

In this exercise, you will use Glide XP to refine a set of ligands taken from the first SP docking run.

1. In the Core tab, select Do not use.

   Core constraints are now turned off.

2. In the Settings tab, under Precision, select XP (extra precision).
3. Under Options, select Refine (do not dock).
4. Select Write XP descriptor information.

   Note: This option requires a special license. If you do not have this license, do not select the option. You will obtain results without the license, but you will not be able to complete the exercise in Section 4.5 on page 33.

5. In the Ligands tab, ensure that File is selected.
6. Click Browse.

   A file selector is displayed.

7. Ensure that Files of type is set to Maestro.
8. Navigate to the `tutorial/structures` directory.
9. Choose `refine_xp_entries.mae.gz`, and click Open.
10. Ensure that the selected Range is from `1` to End (the default).
11. If you did not select Write XP descriptor information, in the Structure output section of the Output tab, select Write pose viewer file (includes receptor; filename will be <job-name>_pv.mae.gz).

    If you did select this option, the Write pose viewer file option is automatically selected.

12. Start the job with the name `factorXa_xp_refine`.

This job may take up to 30 minutes to run on a 2 GHz processor.

Examining Glide Data

In this chapter, Glide results are examined with an emphasis on visual rather than numerical appraisal. The first set of exercises use the Project Table to display the results of a Glide docking job, examine individual ligand poses and their contacts with the input receptor structure. The second set of exercises uses the Glide XP Visualizer panel to display information on the terms in the Glide XP scoring function that contribute to the ligand binding.

If you have not started Maestro, start it now (see Section 1.3).

Before proceeding with the exercises, change the working directory to the `glide` directory. See Section 1.3 on page 3 for instructions on how to do this.

4.1 Importing Pose Data

The first task is to import poses from a pose viewer file into the Maestro project.

1. In the main window, on the toolbar, click the Import structures button.

   

   The Import panel opens.

2. Click on Options.

   The Import Options dialog box opens.

3. Ensure that Maestro is chosen from the Files of type menu.

4. Ensure that Import all structures, Replace Workspace, and Fit to screen following import are all selected.

5. Click Close in the Import Options dialog box.

6. In the Import panel, select the file `factorXa_sp_pv.maegz` and click Open.

The receptor and the ligands are imported as an entry group named factorXa_sp_pv. The receptor is displayed in the Workspace.

4.2 Viewing Poses

The Project Table panel has special options for entry groups that consist of a receptor and a set of ligands. These options are available from the Entry menu, under View Poses, when you have a single entry group selected.

1. Open the Project Table panel.

   You can do this by clicking the Open/Close project table toolbar button.

   

   The entries in the entry group containing the receptor and the poses that you imported should be selected. If not, click in the Row column for the entry group.

2. From the Entry menu, choose View Poses > Setup.

   The receptor is fixed in the Workspace, and the first pose is included in the Workspace. A Mark property is added to the Project Table so you can record any poses that you want to mark as being of special interest. To mark a Workspace entry, type M.

The next few steps change the display so that the ligand fills most of the Workspace.

3. From the Workspace selection toolbar button, choose Molecule.

   

   The A on the button changes to an M, to indicate that molecules are being picked.

4. Click the ligand molecule.

   The atoms in the ligand are marked with pale yellow markers.

5. Click the Fit to screen toolbar button.

   

   The view zooms in so that the ligand fills most of the Workspace.

6. Click in an empty portion of the Workspace to clear the selection.

7. Next turn on Workspace Feedback by going to Display > Show Single-Entry Feedback or by typing S.

By default Workspace Feedback appears in the upper right corner of the Workspace, and includes the entry Title, GlideScore, and EmodelScore.

Now you are ready to view the poses.

8. Press the RIGHT ARROW key.

   The second pose replaces the first in the Workspace. The RIGHT ARROW and LEFT ARROW keys can be used to step through the selected poses.

9. Shift-click the entry for the first pose.

   The first pose is added to the Workspace.

10. Press the RIGHT ARROW key.

    The third pose replaces the first two in the Workspace.

11. Reselect the first pose by clicking its In column.

4.3 Displaying Atoms by Proximity

In this section, you will select a display that includes the ligand and the receptor residues nearest the ligand. This is useful for examining contacts and hydrogen bonds between the ligand and the active site of the receptor.

1. From the Display only selected atoms button menu, choose Molecules.

   

2. Click on an atom in the ligand.

   The ligand molecule is displayed, and all other atoms are undisplayed.

3. From the Display residues within N Å of currently displayed atoms button menu, choose +5 Å.

   

   Only the ligand and the nearby residues are displayed. All residues that do not have any atoms within 5 Å of the ligand are undisplayed. Hiding the residues that do not come into contact with the ligand makes it easier to examine the ligand-receptor interactions.

   You can also open the Atom Selection dialog box from the Display only button menu to pick the ligand and add atoms within a given radius of a particular set of atoms. This approach allows you more flexibility in picking the atoms.

4.4 Visualizing Hydrogen Bonds and Contacts

In this exercise, you will display hydrogen bonds between the ligand and the receptor. (To display hydrogen bonds between any two sets of atoms, use the Atom set 1 and Atom set 2 selection options in the H-Bonds tab.)

1. In the Project Table, choose Entry > View Poses > Display H-bonds.

   Hydrogen bonds to the currently displayed pose are displayed as yellow dashed lines.

   If you want to change the cutoffs for defining hydrogen bonds, choose Entry > View Poses > Define H-bonds, and change the values in the Measurements panel, which is displayed by choosing this item.

   If you want a count of hydrogen bonds between the ligands and the receptor, choose Entry > View Poses > Count H-bonds. An HBond property will be added to the Project Table, and the count may take a few seconds to finish.

2. Use the RIGHT ARROW and LEFT ARROW keys to step through the poses.

   The hydrogen bonds are displayed as each pose is included in the Workspace. Note the difference in hydrogen bonding patterns between the ligands.

In the pose list, 344 Good vdW, 6 Bad vdW, and 1 Ugly vdW contacts are reported for `1dwd`. Even the least-good ligand pose has many more good contacts than bad or ugly ones. The default is to display only Bad or Ugly contacts between the ligand and the receptor. (To display contacts between any two sets of atoms, use the Atom set 1 and Atom set 2 selection options in the Contacts tab.)

3. From the Entry menu, choose View Poses > Display Contacts.

   The contacts are shown as dashed lines connecting Workspace atoms. Ugly contacts are shown in red and Bad contacts are shown in orange. By default, atoms that are hydrogen bonded are not considered to have bad or ugly contacts.

   If you want to change the cutoffs to redefine the distance criteria for Good, Bad, and Ugly contacts, choose Entry > View Poses > Define Contacts, and change the values in the Measurements panel, which is displayed by choosing this item.

   If you want a count of contacts between the ligands and the receptor, choose Entry > View Poses > Count Contacts. The count may take a few seconds to finish, and three new properties are added to the Project Table: Good, Bad, and Ugly.

4. Use the RIGHT ARROW and LEFT ARROW keys to step through the poses.

4.5 Visualizing Glide XP Descriptors

In this exercise, you will use the Glide XP Visualizer to examine the contributions of various terms to the XP scoring function. The terms are given a spatial representation that you can display together with the ligand and the receptor.

Note: This exercise uses the results obtained from the exercise in Section 3.9 on page 26, which requires a special license for XP descriptor generation.

1. Click the Clear Workspace toolbar button.

   

2. From the Applications menu, choose Glide > XP Visualizer.

   The Glide XP Visualizer panel opens.

3. Click Open.

4. Select `factorXa_xp_refine.xpdes` in the file selector that opens, and click Open.

   After a short delay, the receptor and the highest-scoring ligand are displayed in the Workspace, and the table in the panel is filled in. The underlined values indicate that there is a corresponding visualization for this value.

5. Click the PhobEn cell for ligand 16088.

   You might want to deselect Narrow columns to see the entire column heading. This column displays the hydrophobic enclosure rewards. After a few seconds, the naphthalene of the ligand is displayed in ball-and-stick, and the hydrophobic atoms on the protein surrounding this ring are displayed in CPK in gray.

6. Click the PhobEn cell for ligand 612278.

   Note that for this ligand there is only a benzene ring rather than a naphthalene.

7. Click the HBond cell for ligand 16088.

   The hydrogen bonds are displayed as yellow dotted lines and annotated with their lengths.

8. Click the Electro cell for ligand 16088.

   The amidine nitrogens are displayed in ball-and-stick, to indicate their contribution to electrostatic rewards.

9. Click the π Stack cell for ligand 16088.

   The aromatic groups in the protein that contribute to the π stacking reward are displayed in gray in CPK, and the ligand aromatic group is displayed in ball-and-stick.

10. Click the RotPenal cell for ligand 16088.

    The bonds that contribute to the rotatable bond penalty are displayed as tubes.

11. Close the Glide XP Visualizer panel.

12. Clear the Workspace, deleting the scratch entry.

4.6 Finishing the Exercises

Close the scratch project you are working in. Because you have written the output structure files to your directory tree, you do not need to save the scratch project or Workspace structures. Click OK to delete any scratch entries.

From the Maestro menu, choose Quit and click Quit, do not save log file. (For more information about Quit panel options and `maestrolog.cmd` files, click Help instead.)

Getting Help

Schrödinger software is distributed with documentation in PDF format. If the documentation is not installed in `$SCHRODINGER/docs` on a computer that you have access to, you should install it or ask your system administrator to install it.

For help installing and setting up licenses for Schrödinger software and installing documentation, see the *Installation Guide*. For information on running jobs, see the *Job Control Guide*.

Maestro has automatic, context-sensitive help (Auto-Help and Balloon Help, or tooltips), and an online help system. To get help, follow the steps below.

- Check the Auto-Help text box, which is located at the foot of the main window. If help is available for the task you are performing, it is automatically displayed there. Auto-Help contains a single line of information. For more detailed information, use the online help.
- If you want information about a GUI element, such as a button or option, there may be Balloon Help for the item. Pause the cursor over the element. If the Balloon Help does not appear, check that Show Balloon Help is selected in the Maestro menu of the main window. If there is Balloon Help for the element, it appears within a few seconds.
- For information about a panel or the tab that is displayed in a panel, click the Help button in the panel, or press F1. The help topic is displayed in your browser.
- For other information in the online help, open the default help topic by choosing Online Help from the Help menu on the main menu bar or by pressing CTRL+H. This topic is displayed in your browser. You can navigate to topics in the navigation bar.

The Help menu also provides access to the manuals (including a full text search), the FAQ pages, the New Features pages, and several other topics.

If you do not find the information you need in the Maestro help system, check the following sources:

- *Maestro User Manual*, for detailed information on using Maestro
- *Maestro Command Reference Manual*, for information on Maestro commands
- *Maestro Overview*, for an overview of the main features of Maestro
- *Maestro Tutorial*, for a tutorial introduction to basic Maestro features
- *Glide User Manual*, for detailed information on using Glide
- *Protein Preparation Guide*, for information on protein preparation for Glide
- *Impact Command Reference Manual*, for information on Impact commands

- Glide Frequently Asked Questions pages, at https://www.schrodinger.com/Glide_FAQ.html
- Known Issues pages, available on the Support Center.

The manuals are also available in PDF format from the Schrödinger Support Center. Local copies of the FAQs and Known Issues pages can be viewed by opening the file `Suite_2009_Index.html`, which is in the `docs` directory of the software installation, and following the links to the relevant index pages.

Information on available scripts can be found on the Script Center. Information on available software updates can be obtained by choosing Check for Updates from the Maestro menu.

If you have questions that are not answered from any of the above sources, contact Schrödinger using the information below.

E-mail: help@schrodinger.com
USPS: Schrödinger, 101 SW Main Street, Suite 1300, Portland, OR 97204
Phone: (503) 299-1150
Fax: (503) 299-4532
WWW: http://www.schrodinger.com
FTP: ftp://ftp.schrodinger.com

Generally, e-mail correspondence is best because you can send machine output, if necessary. When sending e-mail messages, please include the following information:

- All relevant user input and machine output
- Glide purchaser (company, research institution, or individual)
- Primary Glide user
- Computer platform type
- Operating system with version number
- Glide version number
- Maestro version number
- mmshare version number

On UNIX you can obtain the machine and system information listed above by entering the following command at a shell prompt:

```
$SCHRODINGER/utilities/postmortem
```

This command generates a file named *username*`-`*host*`-schrodinger.tar.gz`, which you should send to help@schrodinger.com. If you have a job that failed, enter the following command:

```
$SCHRODINGER/utilities/postmortem jobid
```

where *jobid* is the job ID of the failed job, which you can find in the Monitor panel. This command archives job information as well as the machine and system information, and includes input and output files (but not structure files). If you have sensitive data in the job launch directory, you should move those files to another location first. The archive is named *jobid*`-archive.tar.gz`, and should be sent to help@schrodinger.com instead.

If Maestro fails, an error report that contains the relevant information is written to the current working directory. The report is named `maestro_error.txt`, and should be sent to help@schrodinger.com. A message giving the location of this file is written to the terminal window.

More information on the `postmortem` command can be found in Appendix A of the *Job Control Guide*.

On Windows, machine and system information is stored on your desktop in the file `schrodinger_machid.txt`. If you have installed software versions for more than one release, there will be multiple copies of this file, named `schrodinger_machid-`*N*`.txt`, where *N* is a number. In this case you should check that you send the correct version of the file (which will usually be the latest version).

If Maestro fails to start, send email to help@schrodinger.com describing the circumstances, and attach the file `maestro_error.txt`. If Maestro fails after startup, attach this file and the file `maestro.EXE.dmp`. These files can be found in the following directory:

```
%USERPROFILE%\Local Settings\Application Data\Schrodinger\appcrash
```

Glossary

Base Name—The name entered in the Base name for grid files text box that is used to write grid files during a grid file calculation, or to find pre-existing grid files during a docking job.

Bounding Box—The green, cube-shaped marker that appears in the Workspace during Glide docking job setup after you select active site residues, coordinates, or a ligand to be used as the box's center. The box represents the space in which ligands are allowed to move during docking. Increasing the size of the bounding box increases the space that can be sampled by the docked ligands, and consequently increases the CPU time required for the calculation.

Contacts—Graphical representations of the van der Waals interactions between the atoms of two or more molecules. Within Maestro, contacts are categorized as "Good," "Bad," and "Ugly." Good contacts are those that have van der Waals radii consistent with the experimentally determined values for the involved atom types. Bad contacts depict those interactions that are experimentally improbable. Ugly contacts represent van der Waals interactions that are disallowed in experimental systems.

Enclosing Box—The purple, cube-shaped marker that appears in the Workspace after you specify active residue sites, coordinates, or a ligand to be used as a bounding box center using the Glide panel. The enclosing box represents the space that any part of any specified ligand can sample during a docking calculation. Compare this with the green *bounding box*, which represents the space that the center of each specified ligand must be confined to during a docking calculation.

Flexible Docking—A job type in which alternate conformations for each ligand are generated during the docking process, and then the interactions between the receptor and the conformers are analyzed. After docking jobs are complete, the conformers, or "poses," are ranked according to their overall interaction with the receptor. The results can be posted to a pose view file, which can be examined using the Glide Pose Viewer panel.

GlideScore—Glide's scoring function (based on ChemScore). GlideScore is used in ranking ligand poses found in docking. In Liaison, GlideScore is used in an alternative binding energy model.

Grid Files—Files written by Glide during grid setup. These files contain data about the properties of the associated receptor and are used during docking.

Ligand Centroid—Used to define the enclosing box center, a ligand centroid is the point whose x, y, and z coordinates are the mean of the minimum and maximum *x*, *y*, and *z* coordinates of all the atoms in the ligand.

Pose Viewer Panel—An analysis tool that displays the results of Glide docking jobs. These results, which are recorded in a pose view file, include the ligand name, pose number, overall score, number of contacts, and other data. The poses within the file are arranged in the list according to score: ligands with the most energetically favorable interactions with the receptor appear at the beginning, and ligands with less favorable interactions appear near the end. The Glide Pose Viewer panel can also be used to visualize contacts and hydrogen bonds between ligand and receptor molecules, or to write structure files containing one or more ligand poses.

Reference Ligand—A user-specified structure whose ligand/receptor docking score will be compared with all other docked ligands.

Rigid Docking—A job type in which only supplied conformations of the specified ligands will be docked, scored, and displayed in a pose view file. This job type is useful if you have already performed a conformational search on the ligands that you want to dock.

120 West 45th Street, 29th Floor
New York, NY 10036

101 SW Main Street, Suite 1300
Portland, OR 97204

8910 University Center Lane, Suite 270
San Diego, CA 92122

Zeppelinstraße 13
81669 München, Germany

Dynamostraße 13
68165 Mannheim, Germany

Quatro House, Frimley Road
Camberley GU16 7ER, United Kingdom

SCHRÖDINGER®

Docking with GLIDE Summary of Main Steps

Project → New (give a name)
Table → Import → Structures (select PDB file)

By using DUPLICATE: creation of new entries containing:
- macromolecule + cofactor (if any) + ions + water molecules to be kept + …

Icons on the left:

 Color by Chain Name (Palette icon left bottom)

 Delete chain B (red)

 Center view

 Color by Element

 Ribbon | Display side chains

 Hydrogen bonds visualization

1) Macromolecule preparation
WORKFLOWS: - Protein Preparation Wizard

→ Preprocess:
- Assign bond orders
- Add hydrogens
- Delete waters
…etc.
→ Optimize + Minimize

ALWAYS: check Job Monitor Status.

2) Ligand preparation

- Change bond order | → ||
- Edit → Atom Properties → Atom type
- Add hydrogens: +H
- Edit → Clean-up geometry
- LigPrep: generate possible states at target pH 7 (Ionizer) → Start

3) GLIDE -> Receptor Grid generation

Site → Centroid of selected residues
Constraints → H-bond
Rotatable groups

→ Start

4) GLIDE → Ligand Docking → Receptor Grid name

- SP
- Dock flexibly
- Ligands: workspace
- Constraints
- Output: number poses per ligand, write per-residue interactions scores, compute RMSD, write report
- Start: incorporate new entries as a new group

5) Analysis: Open poseviewer file name_pv.maegz)

 View table and setup Pose display: Entry → View Poses → Setup

Compare docking results with native solutions